

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2002, 05:37:35 ; Search time 5265 seconds  
(without alignments)  
16995.607 Million cell updates/sec

Title: US-09-730-559B-1  
Perfect score: 4276  
Sequence: I ttctaccgtttttccctgc.....atcagaaaaaaaaa 4276

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg0.inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	3690.4	86.3	167659	2	AC023404	AC023404 Homo sapi
2	3690.4	86.3	175793	2	AC093300	AC093300 Homo sapi
3	3690.4	86.3	180130	2	AC093275	AC093275 Homo sapi
c	42480.4	58.0	157320	2	AC027609	AC027609 Homo sapi
5	912.6	21.3	6380	9	AB051460	AB051460 Homo sapi
6	467.4	10.9	212601	2	AL662793	AL662793 Mus muscu
c	467.4	10.9	249657	2	AL665946	AL665946 Mus muscu
8	430	10.1	5715	9	AB023157	AB023157 Homo sapi
9	336.4	9.3	71171	2	AC100917	AC100917 Mus muscu
10	304.6	7.1	3114	3	AY052043	AY052043 Drosophil
c	304.6	7.1	103054	3	AC014803	AC014803 Drosophil
c	304.6	7.1	181720	3	AC010114	AC010114 Drosophil
c	304.6	7.1	287018	3	AE003553	AE003553 Drosophil
c	240.8	5.6	205307	2	AC009968	AC009968 Homo sapi
c	238	5.6	37037	9	AC004973	AC004973 Homo sapi
c	238	5.6	126000	9	AP000744	AP000744 Homo sapi
c	238	5.6	135038	9	HUMYWXD703	L78810 Homo sapien
c	237	5.5	113687	9	AC069281	AP002354 Homo sapi
c	237	5.5	179245	9	HUAC002492	AC069281 Homo sapi
c	231.4	5.4	203959	2	AC064827	AC064827 Homo sapi
c	230.4	5.4	133475	9	AC006254	AC006254 Homo sapi
c	230.2	5.4	148300	2	AC091489	AC091489 Homo sapi
c	230.2	5.4	181532	9	CNS01DW2	AL136298 Human chr
c	230.2	5.4	187696	2	AC008785	AC008785 Homo sapi
c	229.4	5.4	155691	9	AC016868	AC016868 Homo sapi
c	229.4	5.4	214872	2	AC009720	AC009720 Homo sapi
c	228.6	5.3	166889	2	AC027008	AC027008 Homo sapi
c	228.6	5.3	167412	2	AL356352	AL356352 Homo sapi
c	228.6	5.3	169963	9	AL158207	AL158207 Human DNA
c	228.4	5.3	241288	2	AC011499	AC011499 Homo sapi
c	228.4	5.3	147224	2	AP003779	AP003779 Homo sapi
c	228.4	5.3	161262	2	AC087464	AC087464 Homo sapi
c	228.4	5.3	171285	9	AL353662	AL353662 Human DNA
c	228.4	5.3	208729	2	AC012052	AC012052 Homo sapi
c	228.4	5.3	215177	2	AC011667	AC011667 Homo sapi
c	228	5.3	162589	2	AC022076	AC022076 Homo sapi
c	228	5.3	180465	9	AC074011	AC074011 Homo sapi
c	228	5.3	196759	2	AC092164	AC092164 Homo sapi
c	227.8	5.3	84388	9	AL589984	AL589984 Human DNA
c	227.4	5.3	159634	2	AC010193	AC010193 Homo sapi
c	227.2	5.3	163891	9	AC092647	AC092647 Homo sapi
c	227	5.3	172789	2	AC104020	AC104020 Homo sapi

ALIGNMENTS

RESULT	1	AC023404	167659 bp	DNA	linear	HTG 01-MAR-2000
LOCUS	AC023404	Homo sapiens clone RP11-619L12, WORKING DRAFT SEQUENCE, 10				
DEFINITION	1	unordered pieces.				
ACCESSION	AC023404	GI:7139760				
VERSION	AC023404.2	HTGS_PHASE1; HTGS_DRAFT.				
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N., Anderson S., Baldwin J., Barna N., Bada F., Bozulski L., Boukhgalter B., Brown A., Burkett G., Campopiano A., Castle A., Choepey Y., Colangelo M., Collins S., Collipalano A., Cooke P., DeArrellano K., Dewar K., Dodge S., Domino M., Doyle M., Fenestor J., Ferreira P., Fitzhugh W., Forrest C., Gage D.,					
JOURNAL	Unpublished					
REFERENCE	2	(bases 1 to 167659)				
AUTHORS	Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N., Anderson S., Baldwin J., Barna N., Bada F., Bozulski L., Boukhgalter B., Brown A., Burkett G., Campopiano A., Castle A., Choepey Y., Colangelo M., Collins S., Collipalano A., Cooke P., DeArrellano K., Dewar K., Dodge S., Domino M., Doyle M., Fenestor J., Ferreira P., Fitzhugh W., Forrest C., Gage D.,					





Db	136712	ACATGTATGCTAAGCATGTCGTCCCTAGCTAGCTAGCAAAATGTAATCTCTATTTTGATT		AC093300.1	GI:15193434
QY	3349	gcagggtgaaggttaagccatatctcttgatgatcagctgtgtgatgaatgcaggggcc		HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.	
Db	136772	GCAGGTGAAGTTAAGCCATATCTCTGGATGATACAGCTGTGTGATGAATGTCAGGGGGC		Human.	
QY	3409	ccgtgtgggggaaattgctccattttctgtgctaattgttacctgtctcagtatata		Homo sapiens	
Db	136832	CCGTTGGGGGGAAATTTGCTCCATTTTCTGTGCTAATGTTACCTGCTGCAGTATTA		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
QY	3469	ctgtgaatatgtcggctgctatccattctgtcgtgcagggaattccacaagccctc		Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
Db	136892	CTGTGAATATTGTGGGCTGCTATCCATTTCTGTGTGGCAGGAATCCACAAGCCCT		DOE Joint Genome Institute.	
QY	3529	ggtgaagaagcggtgacgcgcctcgcatatttccattccctggaactaaagataac		Sequencing of Human Chromosome 5	
Db	136952	GGTGAAGGAAGGCGGTGACGCGCCTCGGCATATTTTCATCCGCTGGAATGAAGATAAC		Unpublished	
QY	3589	tgcagtgtctcattttcaggcctcagaataagtgacactctctgttcatctgacccctc		2 (bases 1 to 175793)	
Db	137012	TGCAGTGTCTCATTTTCAGGCTCAGAAATAGTGCACCTCTCTGTTTCATTTGACCCCTTC		DOE Joint Genome Institute.	
QY	3649	ctcaacctcttcacgtggtcattctctgttagcagctctgtaacttaactatagtataa		Direct Submission	
Db	137072	CTCAACCTCTTCACGCTGGCATGCTCCTTTTGTAGCAGCTCTGTAACCTTAACATAGTATAA		Submitted (16-AUG-2001)	
QY	3709	tgaagaagaatgacctataatagggtgtttttagattctgtgacactgcaaaacaatat		Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
Db	137132	TGAAAGAATAGACTATAATATAGGTGTTTTGTAGATCTCTGTGTGCACATGCAACAATAT		-----Genome Center	
QY	3769	gaacctctttctgattgcatcggtgtgcatggaagtttattctctgttttctgtg		Center: Joint Genome Institute	
Db	137192	GAACTCTTTTTCGATGCGATCGGGTTCATGGAAGTTTATTCTCTGTTTCTGG		Center Code: JGI	
QY	3829	aaacaagagatccaaactctgcacaactttcttagaggagagagagaatataaa		Web site: http://www.jgi.doe.gov	
Db	137252	AAACAAAGAGATCCAACTTCCTGCAACATTTCTTAGAGGAGAGAGAGAATATATAA		-----	
QY	3889	agagaatatgaaatagatgttttgggtttttaaattattgttaataataataac		Project Information	
Db	137312	AGAGAAATGAACAANTAGATATTTTGGGTTTTTAATTAATTTATTGTTAAATAATAAC		Center Project Name: 621412	
QY	3949	ataagaataacttttataaaataaccatgcaacaataaacactatcgctctatctgaca		Center clone name: RPCI-11_560A7	
Db	137372	ATATAAGAATACTTTATTAAATAACCATGCAACAATAAACACTATCGCTATCTGACA		-----	
QY	4009	gttttccccccagggaagtcgttttgccttttctccttttcttttttttttttttctcttt		Summary Statistics	
Db	137432	GTTTTTCCCCCAGGAAGTCGTTTTGCGCTTTTCTCTTTTCTTTTCTTTTCTTCACTT		Consensus quality: 162106 bases at least Q40	
QY	4069	tttgtctctctctttttccatcccttttttaatttttaacagcaatggaggaggtta		Consensus quality: 170396 bases at least Q30	
Db	137492	TTTTGCTCTCTCTCTTTTCCATCCCTTTTAAATTTTAAACAGCAATGGAGGAAGTTA		Consensus quality: 171515 bases at least Q20	
QY	4129	acaatttttaaggaagagcagctgttagagcaacaataatgcaataagcagactgagcagc		Estimated insert size: 104030; agarose-fp estimation	
Db	137552	ACAAATTTTAAATGAAAGAGCATGTTAGAGCAACAATAATGCAAGCAAGACTGAGCAGC		Quality coverage: 6.21 in Q20 bases; sum-of-contigs estimation	
QY	4189	attataatttaattcaagggttttgaggctgaaacataaattcattatccctcaaaagtt		Quality coverage: 3.71 in Q20 bases; sum-of-contigs estimation.	
Db	137612	ATTATAATTAATTTTCAGGGTTTGGAGGCTGAACATAATTTCTATTATCCCTCAAAAAGTT		* NOTE: This is a 'working draft' sequence. It currently	
QY	4249	acaccacatcagaaaaaataaaaaa 4276		* consists of 19 contigs. The true order of the pieces	
Db	137672	ACCACCACATCAGAAAAATAAAAAA 137699		* is not known and their order in this sequence record is	
RESULT	2			* arbitrary. Gaps between the contigs are represented as	
AC093300/c				* runs of N, but the exact sizes of the gaps are unknown.	
LOCUS				* This record will be updated with the finished sequence	
DEFINITION				* as soon as it is available and the accession number will	
				* be preserved.	
				* 1749: contig of 1749 bp in length	
				* 1750: gap of unknown length	
				* 1850: contig of 1494 bp in length	
				* 3343: gap of unknown length	
				* 3443: contig of 1498 bp in length	
				* 4941: gap of unknown length	
				* 5041: gap of unknown length	
				* 5042: contig of 2663 bp in length	
				* 7705: gap of unknown length	
				* 7805: contig of 1557 bp in length	
				* 9362: gap of unknown length	
				* 9461: contig of 2507 bp in length	
				* 11968: gap of unknown length	
				* 12068: contig of 3643 bp in length	
				* 15711: gap of unknown length	
				* 15712: contig of 4585 bp in length	
				* 15812: gap of unknown length	
				* 20397: contig of 3939 bp in length	
				* 20497: gap of unknown length	
				* 24435: gap of unknown length	
				* 24536: contig of 7595 bp in length	
				* 32130: gap of unknown length	
				* 32230: contig of 8082 bp in length	
				* 40312: gap of unknown length	
				* 40413: contig of 8248 bp in length	
				* 48660: gap of unknown length	
				* 48661: contig of 10616 bp in length	
				* 48761: gap of unknown length	
				* 59376: contig of 11137 bp in length	
				* 59377: gap of unknown length	
				* 70613: contig of 18333 bp in length	
				* 70714: gap of unknown length	
				* 89047: contig of 18317 bp in length	
				* 89147: gap of unknown length	
				* 107463: contig of 18317 bp in length	
				* 107464: gap of unknown length	











```

Db 19182 TTTTCTCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAATTA 19241
QY 4129 acaatttttaagaaagagcatgttagagcaacaataagcaagactgagcagc 4188
Db 19242 ACAATTTTAAATGGAAGAGCATGTTAGAGCAAAACAATGATAAGCAAGACTGAGCAGC 19301
QY 4189 attataaatttttcagggttttgagcgtgaacataatttcattccctcaaaaagtt 4248
Db 19302 ATTATAATTAATTTTCAGGGTTTGGCGCTGAACATAATTTCATTCCTCAAAAAGTT 19361
QY 4249 accaccacatcagaaaaaataaaaaa 4276
Db 19362 ACCACCACATCAGAAAAATAAAAAA 19389

RESULT 4
AC027609/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone RP11-80E21, WORKING DRAFT SEQUENCE,
34 unordered pieces.
ACCESSION AC027609
VERSION AC027609.3 GI:9958118
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157320)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 157320)
Waterston,R.H.
Direct Submission
Submitted (30-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637328.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH080E21
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Consensus program: Phrap; version 0.990319
Consensus quality: 138437 bases at least Q40
Consensus quality: 143665 bases at least Q30
Consensus quality: 145934 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 153359; sum-of-contigs
Quality coverage: 3.44 in Q20 bases; agarose-fp
Quality coverage: 3.89 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1657: contig of 1657 bp in length
* 1658 1757: gap of unknown length
* 1758 3454: contig of 1697 bp in length
* 3455 3554: gap of unknown length
* 3555 5136: contig of 1582 bp in length

* * * * *
5137 5236: gap of unknown length
5237 7177: contig of 1941 bp in length
7178 7277: gap of unknown length
7279 8325: contig of 1048 bp in length
8326 8426 9445: gap of unknown length
8426 9445: contig of 1020 bp in length
9446 10206: gap of unknown length
9446 10206: contig of 661 bp in length
10207 10307: gap of unknown length
10307 12221: contig of 1915 bp in length
12222 12321: gap of unknown length
12322 14246 14345: contig of 1924 bp in length
14246 16858: gap of unknown length
16859 16959: gap of unknown length
16959 19404: contig of 2446 bp in length
19405 19504: gap of unknown length
19505 21718: contig of 2214 bp in length
21719 21818: gap of unknown length
21819 25226: contig of 3408 bp in length
25227 25326: gap of unknown length
25327 28747: contig of 3421 bp in length
28748 28848: gap of unknown length
28848 32938: contig of 4091 bp in length
32939 36247: contig of 3209 bp in length
36247 36347: gap of unknown length
36348 39886: contig of 3539 bp in length
39887 43666: contig of 3680 bp in length
43667 43766: gap of unknown length
43767 47542: contig of 3776 bp in length
47543 52042: contig of 4400 bp in length
52043 52143 55670: contig of 3528 bp in length
55671 61439: contig of 5669 bp in length
61440 61539: gap of unknown length
61540 65392: contig of 3853 bp in length
65393 65493 69190: contig of 3698 bp in length
69191 76123: contig of 6833 bp in length
76124 76223: gap of unknown length
76224 82113: contig of 5890 bp in length
82114 88806: contig of 6593 bp in length
88807 88907 95966: contig of 7060 bp in length
95967 96066: gap of unknown length
96067 102190: contig of 6124 bp in length
102191 102290: gap of unknown length
102291 112232: contig of 9942 bp in length
112233 112332: gap of unknown length
112333 122699: contig of 10367 bp in length
122700 122799: gap of unknown length
122800 128543: contig of 6744 bp in length
128544 129644 143636: contig of 13993 bp in length
143637 143736: gap of unknown length
143737 157320: contig of 13584 bp in length.

* * * * *
Location/Qualifiers
1..157320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-80E21"
1..1657
/note="assembly_name:Contig14"
1758..3454
/note="assembly_name:Contig15"
3555..5136

FEATURES
source
misc_feature
misc_feature
misc_feature
```

misc\_feature /note="assembly\_name:Contig17"  
5237. .7177  
misc\_feature /note="assembly\_name:Contig18"  
7278. .8325  
misc\_feature /note="assembly\_name:Contig19"  
8426. .9445  
clone\_end:SP6  
vector\_side:left  
misc\_feature 9546. .10206  
misc\_feature /note="assembly\_name:Contig10"  
10307. .12221  
misc\_feature /note="assembly\_name:Contig21"  
12322. .14245  
misc\_feature /note="assembly\_name:Contig22"  
14346. .16858  
misc\_feature /note="assembly\_name:Contig23"  
16959. .19404  
misc\_feature /note="assembly\_name:Contig24"  
19505. .21718  
misc\_feature /note="assembly\_name:Contig25"  
21819. .25226  
misc\_feature /note="assembly\_name:Contig26"  
25327. .28747  
misc\_feature /note="assembly\_name:Contig27"  
28848. .32938  
misc\_feature /note="assembly\_name:Contig28"  
33039. .36247  
misc\_feature /note="assembly\_name:Contig29"  
36348. .39886  
misc\_feature /note="assembly\_name:Contig30"  
39987. .43666  
misc\_feature /note="assembly\_name:Contig31"  
43767. .47542  
misc\_feature /note="assembly\_name:Contig32"  
47643. .52042  
misc\_feature /note="assembly\_name:Contig33"  
52143. .55670  
misc\_feature /note="assembly\_name:Contig34"  
55771. .61439  
misc\_feature /note="assembly\_name:Contig35"  
61540. .65392  
misc\_feature /note="assembly\_name:Contig36"  
65493. .69190  
misc\_feature /note="assembly\_name:Contig37"  
69291. .76123  
misc\_feature /note="assembly\_name:Contig38"  
76224. .82113  
misc\_feature /note="assembly\_name:Contig39"  
82214. .88806  
misc\_feature /note="assembly\_name:Contig40"  
88907. .95966  
misc\_feature /note="assembly\_name:Contig41"  
96067. .102190  
misc\_feature /note="assembly\_name:Contig42"  
102291. .112232  
misc\_feature /note="assembly\_name:Contig43"  
112333. .122699  
misc\_feature /note="assembly\_name:Contig44"  
122800. .129543  
misc\_feature /note="assembly\_name:Contig45"  
129644. .143636  
misc\_feature /note="assembly\_name:Contig46"  
143737. .157320  
misc\_feature /note="assembly\_name:Contig47"  
3317 others  
BASE COUNT 45127 a 32028 c 32232 g 44616 t 3317 others  
ORIGIN

Query Match 58.0%; Score 2480.4; DB 2; Length 157320;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2509; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

Qy 1756 aagagaaaagtatctgtcttattctaagcatgggggagggacaagattagtagttaaca 1815  
Db 43666 AAGAGAAAAGTATCTGTCTATCTAAGCATGGGGAGGGACAAGATTAGTATGTTAACA 43607  
Qy 1816 tgctactttgtttgtttgagatggagctctctctcogtccaccaggtcgagtgagtg 1875  
Db 43606 TGCCTACTTTGTTGTTGATGGAGTCTCTCTCCGTACCCAGGCTGGAGTGCAGTGG 43547  
Qy 1876 tacagttcagctcaactccaactctgcctccgggttcaagtgtattcttcctgccttagc 1935  
Db 43546 TACAGTCTCAGCTCACCAACCTCTGCCCTCCCGGGTTCAAGTGATTCTCCGCTTAGC 43487  
Qy 1936 ctcccgagtagtggaattacagggcatataaccaccatgcccaacaaatgtttgtattttt 1995  
Db 43486 CTC CGAGTAGTGCAATTACAGGCATATACCAACATATGCCAACAATAATGTTGTATTTT 43427  
Qy 1996 agtgagacaggggtttcacogtgttggtcaaggccaggtttcaaaactcctgacctcaaggga 2055  
Db 43426 AGTGGAGACAGGGTTTCAACGTGTTGGTCAGGCCAGTTTCAAACTCCTGACCTCAAGGGA 43367  
Qy 2056 tccacctgcctcacccctcaaaagtgtgggattacagggcatgagccaccacctagcct 2115  
Db 43366 TCCACCTGCCCTCACCCCTCAAAAGTGTGGGATTACAGGCATGAGCCACCCACCATGCCT 43307  
Qy 2116 ggcctacttggttttttatgcacactaaaaataacctacatctcactgccttattccaac 2175  
Db 43306 GGCCTACTTGGTTTTTATGCACACTAAAAATACCTACATCTCACTGCTTATTCCCAAC 43247  
Qy 2176 ataagttcagagctgtggatgggtcattagaaatcagactgaattgtgtctctcg 2235  
Db 43246 ATAAGTTTCAGAGCTGTGGGATTGGTTCATTAGAAATTCAGACTGAATTTGTGTTCTCTCG 43187  
Qy 2236 caatgaatcctttggccagtggttcattgtcactctctgtacattatggagcgacctagag 2295  
Db 43186 CAATGAATCCTTTGCCAGTGTTCATGTCACTCTCTAGACATATGAGAGCGCTAGAG 43127  
Qy 2296 gccagaagccagctgcctccttatgcctctctctcctgggttcctgacactctcttc 2355  
Db 43126 GCCAGAAGCCAGTGCTCTCTTATGCCCTCTCTCTGGGCTTCGTGACACTCTTCTTC 43067  
Qy 2356 tcctttgtacttttttttttagttaaaaaatTTTTTTtagagggagggtctcactct 2415  
Db 43066 TCCTTTTGTACTTTTATTTTTTTAGTTAAAAAATTTTTTTAGAGGGAGGGTCTCACTCT 43007  
Qy 2416 gtaccagagctggagcaagaatacaaatcagactcactgcattcttctctcttttg 2475  
Db 43006 GTCACCCAGGCTGGAGCACGAATCAATCATGCTCACTGCATGCTGCTCTCTCTTTTG 42947  
Qy 2476 ttcatggctaattgttgctcaggttctcttgcagagctgggtggcaccagtgctggtgac 2535  
Db 42946 TTCATGGCTAATCTTGGTCAGGATTCCTTGTCAAGCTGGGTGGCACCAGTCTGGTGAC 42887  
Qy 2536 agcctgtgtaagggagtttcagccatgaatctctccagactaaaaataaccagctcttt 2595  
Db 42886 AGCCTGTGTAAGGGAGTTTTCAGCCATGAATCTCTCCAGACTAAAAAATAACCAAGTCTTT 42827  
Qy 2596 tctagctgtagaatataaccaggtgactgttaagtctgaaggttcacatgcaggt 2655  
Db 42826 TCTAGCTGATGAATTAATAACCAAGGTGACTGTGTAAGCTTTGAAAGGTTCAATGACAGGT 42767  
Qy 2656 tggccgatagaacgctggaaacagggccagtttttagaaattccactctgacttttagaetc 2715  
Db 42766 TGGCCGATAGAACGCTGGAACAGGCCAGTTTTTAGAAAATTCACCTCTGACTTTTAAAGTCTC 42707  
Qy 2716 aggtgaaccattctactgaaagaaacaaagcagggtttttagactgtgaatccctatgac 2775  
Db 42706 AGGTGAACCAATCTTACTGAGAAAGAACAAAGCAGGGTTTTTAGCTGTGAATCTTATGCC 42647  
Qy 2776 tgcac-ttttttttttttaacagaggttccaggtttgtgattataaccaccaactgtgt 2834  
Db 42646 TGCATCTTTTTTTTTTTTAAACAGAGTTCAGAGTTTCCAGGTTTGTGATTATTAACCAACATGTGT 42587  
Qy 2835 acactataatagaaaccacagagcagggttttttaacagagctcagaaatcttctgtgacgca 2894



```

|||||
Db 42586 ACACATATAAAGAACACGAGCGCTTTTACGACACCTCAGAACTCTGTGACGCA 42527
QY 2895 gtagcagcgcattctcacaccgacttgattgaagtgcagttgtggaacttgatc 2954
Db 42526 GTAGTCAGGCATCTTCACACCAACTTGAATATGAAGTGCAGTTGTGTGGAACTTGGATC 42467
QY 2955 atcttagtgatttggtttaattatgattccaatatgacaaaaatccagatccactaa 3014
Db 42466 ATCTTAGTGTGATTTTGTAAATATGATTCACATATGACAAAAATCCAGATCCACTAA 42407
QY 3015 ttaaaatgaggggttatgtctatgaataatctctgtgggtttaaotcctaaacatctta 3074
Db 42406 TTAATAATGAGGGTTATGCTATGAAATATCTCCTGCGGTTAATCATAAACAATCTA 42347
QY 3075 gtctaaacagttgggttcacttcattgatgtctgctcaaatccttttctttaaagatg 3134
Db 42346 GTCTAAACAGTTGGCTTCATCTCATGATGTCTGCTCAAAATCCTTTTCTCTTAAAGGATG 42287
QY 3135 ttatttaataagaataaaatgtaaaatgatagataataaagccttaactagttctta 3194
Db 42286 TTTATTTAATAAG-AAAAAATGTAATAATGATAGATAATAAAGCCTTACTAGGTCTCTTA 42228
QY 3195 aaagatgaactatccatatcttcagtaaatgaataattagtcctctctcttgggcacctt 3254
Db 42227 AAGATGAACATATCCATATTTTCAGTAATGAATAATTAGTCTCTCTTTGGGCACCTT 42168
QY 3255 ggaacagattcattcagatagtggtggaatgtacatgtatgtaacatgtctggcct 3314
Db 42167 GGAACAGATTTCATCAGATAGTGGTGGAAATGTACATGTATGTTAAGCATTTGGTGCT 42108
QY 3315 agtctactgaataatgaactcttatttttaattgcaagtgaaagttaagccatatgtct 3374
Db 42107 AGTCACTGAAAATGTAAACTCTTATTTTGATTGAGGTGGAAGTTAAGCCATAATGCT 42048
QY 3375 tggatgatcagctgtgtgatgaatgtcagggggcccggttggggggaatttgcctcat 3434
Db 42047 TGGATGATCAGCTGTGTGATGAATGTCAGGGGCCCGTGTGGGGGAAATTTGCTCCAT 41988
QY 3435 tttctgtgctaattcactgctgcagttactgtgaattactgtgaattactgctggcgtctatcc 3494
Db 41987 TTTTCTGTGCTAATGTACCTGTCTGCAGTATTAAGTGTGAATATGCTGGGCTGTATCC 41928
QY 3495 attctgtctggcaggaattccaaagccctgtgtaagaaagcgtgacccctc 3554
Db 41927 ATTCCTGCTGTGCGAGGAATTTCCACAAGCCCTGTTGTAAGGAAGCGGTGACTCCCTC 41868
QY 3555 ggcattttcattccgctggaactaaaggataaactgcagtgctcattttcagggcctcaga 3614
Db 41867 GGCATATTTTCATCCGCTGGAACCTAAAGGATAAATGCAGTGTCTATTTTCAGGCCCTCAGA 41808
QY 3615 ataagtgcactctctgttcatctgcacccctctcctaaccctctcagcgtggcatgtcc 3674
Db 41807 ATAAGTGCATCTCTCTGTTCATTTCTGACCCCTCTCTCAACCTCTTTCAGCTGGCATGTCC 41748
QY 3675 tttgtagcagctgtgaacttaactatagataatgaagaagatcacctataatagtt 3734
Db 41747 TTTTGTAGCAGCTGTGAACCTTAACATAGTATAATGAAGAAGTAAGACCTATATAATAGGT 41688
QY 3735 gttttgtagattctgttgcactgcaaaaatataatgaactccttttctgtattgccaatgg 3794
Db 41687 GTTTTGTAGATCTCTGTGTCACCTGCAACAAATATGAACCTCTTTTCTGTATTGCCATCGG 41628
QY 3795 gttgcagtgaagttttattctctgttctgttggaaccaaagagatccaaactctctgc 3854
Db 41627 GTTTCATGGAAGTTTATTTCTCTTGTGTTTTCGTTGGAACCAAGAGGATCCAAACTTCTGTC 41568
QY 3855 aacattttctagagagagagaaataatgaagaagaataatgaacaatacagattttt 3914
Db 41567 AACATTTCTTAGAGAGAGAGAGAAATATTAAGAGAGAAATGAACAATAGAGTATTTT 41508
QY 3915 ggggttttaataattattgttaataataacatatagaatactttttataaataaa 3974
|||||
Db 41507 GGGTTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 41448
QY 3975 ccatgcaacaataaactatcggtctatctgacagtttttccccagggaagtgttttg 4034
Db 41447 CCATGCAACAATAAACAATATCGGTCTATCTGACAGTTCCTCCCGGAGGAGTGTGTTG 41388
QY 4035 ccttttccctctcttttttttttttttttttttttttttttttttttttttttttttttttt 4094
Db 41387 CCTTTCCTTCTCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTC 41328
QY 4095 tttttaaatttttaacagcaatggaggaatttaacaattttttaaatttttaatttttcagggtttt 4154
Db 41327 TTTTAAATTTTAAACAGCAATGGAGAAATTAACAATTTTAAATGGAAGAGCATGTT 41268
QY 4155 agggcaaaaataatgcataaagcagactgagcagcattataatttaatttttcagggtttt 4214
Db 41267 AGAGCAAAACAATAATGCAATAAGCAAGACTGAGCAGCATATATAATTAATTTTCAGGGTTTGA 41208
QY 4215 ggctgaacataaatttcattatccctcaaaaagtaccaccacatcagaaaaaataaaaaa 4274
Db 41207 GGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCACCACATCAGAAAAATAAAAA 41148
QY 4275 aa 4276
Db 41147 AA 41146
|||||
RESULT 5
AB051460
LOCUS AB051460 6380 bp mRNA linear PRI 07-FEB-2001
DEFINITION Homo sapiens mRNA for KIAA1673 protein, partial cds.
ACCESSION AB051460
VERSION AB051460.1 GI:12697890
KEYWORDS
SOURCE Homo sapiens cDNA to mRNA, clone:fg00690.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Nagase,T., Kikuno,R., Hattori,A., Kondo,Y., Okumura,K. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.
XIX. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
DNA Res. 7 (6), 347-355 (2000)
MEDLINE 21082932
REFERENCE 2 (bases 1 to 6380)
Ohara,O., Nagase,T. and Kikuno,R.
Direct Submission
Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdna1nfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)
FEATURES
source
1. .6380
location=Homo sapiens
db_xref="taxon.9606"
clone="fg00690"
note="vector:pBluescriptII SK plus"
69. .2207
gene="KIAA1673"
<69. .2207
/gene="KIAA1673"
note="Start codon is not identified."
/codon_start=1
/product="KIAA1673 protein"
/protein_id="BAB21764.1"
/db_xref="GI:12697891"
/translation="MGDYGFVLVQSNLTKNSAPVRFPHLOPHHONATPSPAF
INNTAANGSAGSAWFLPAPATHNQDEILGSEKAKSQOQQDPLEKQLSPGQ
EAGLLPETERAKSENGNGEKIRIESPVLTFDYEATGLGTSTQPLTS
SASLTGFSNWSAIAFSSSTIINEDASFFHOGVPAASANNALLQNFPHVSPGF
GGSFSPQIGLSQHPHPHFQHHHSQHOQRRSPASPHPPPTHRNAAPNPLHAN

```



BASE COUNT	1895 a	1163 c	1217 g	2105 t
ORIGIN				

[illegible]

```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information -----
Center project name: DM186D6
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 211960 bases at least Q40
Consensus quality: 212056 bases at least Q30
Consensus quality: 212130 bases at least Q20
Insert size: 212301; sum-of-contigs
Insert size: 209170; 13.0% error; agarose-fp
Quality coverage: 13.41x in Q20 bases; sum-of-contigs Quality
coverage: 14.00x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```



```
fragment_chain:1"
79967. .85184
/note="assembly_fragment:00077
fragment_chain:1"
85285. .103614
/note="assembly_fragment:00749
fragment_chain:1"
103715. .132151
/note="assembly_fragment:03729
fragment_chain:1"
132252. .147168
/note="assembly_fragment:03424
fragment_chain:1"
147269. .152571
/note="assembly_fragment:04053
fragment_chain:1"
152672. .177942
/note="assembly_fragment:04381
fragment_chain:1"
178043. .190101
/note="assembly_fragment:00306
fragment_chain:1"
190202. .217559
/note="assembly_fragment:04881
fragment_chain:2"
217660. .249657
/note="assembly_fragment:03901
fragment_chain:2"
BASE COUNT 70686 a 54329 c 54183 g 68743 t 1716 others
ORIGIN

Query Match 10.9%; Score 467.4; DB 2; Length 249657;
Best Local Similarity 78.8%; Pred. No. 2.9e-90;
Matches 738; Conservative 0; Mismatches 146; Indels 53; Gaps 13;

QY 3341 ttttgattgcagtggaagtaagccatcatgtcttgggatgcagctgtgtgatgaatgt 3400
Db 210842 TCTCCAATGCAGTGGAGGTTAAGCCATATGTTGGATACCAGCTGTGTGATGAATGT 210783

QY 3401 cagggggccgtgtgggggaaattgctccattttctgtctgaatgttaccctgtcgt 3460
Db 210782 CAAGGGCCCGTGTGGGGGAAATTTGCTCCATTTTCTGTGTAATGTTACCTGTCTG 210723

QY 3461 cagtattactgtgaattgtgggctgctatccattctgctgctggcagggaattccac 3520
Db 210722 CAGTATTACTGTGAATATTGCTGGGCTGCTATTCTACTCTCGTGGCAGAGATTCCAC 210663

QY 3521 aagccctgtggaagagggcgtgaccgcctcgccatattcattccgctggaact-- 3578
Db 210662 AAGCCCGTGGTGAAGGAAGGTGGTGACCGCCTCGGCATATTTTCATTCGCTGGAAC 210603

QY 3579 --aaggataactgcagtgctcattttcaggccctcagaataagtcgactcttctgtcat 3636
Db 210602 AGAATGAATGACTCAGAGACTCATTTCAGGCGCTCAGAAGAAGTGCCTCTTTGTTAAT 210543

QY 3637 tctgacctctctcaacctctcaactgctgcatgctcttctgttagcagctgtgaactta 3696
Db 210542 TC-CAACTCTTCTCAGCCTCTTTCACATGGCATGTCC-TGCACAGCAGTGTGTAACTA 210486

QY 3697 actatagataatgaaagaatgacctataataggtgtttttagattcttctgtcact 3756
Db 210485 TCAGTGGTAGAATGAAAGAACAAACATACATAGGTATTTTGTAGACTCTCGTGTAC 210426

QY 3757 tcaacaataatgaactcct--ttttcgattgcatcgggttcgcatggaagttttt-att 3813
Db 210425 TGCACAACAGTAGTTCAACTCCATTTCACATTGCCAATATGTCATGTCAGGTTTAAAT 210366

QY 3814 ctctgttttctgctgaaacccagaggtaccacactctcctgcacatcttcttagaggaga 3873
Db 210365 CACTGTGTTTCTGCGGCCCAAGAGGATCTAACTGCTGTAGCATTTTCTTAAG---- 210308

QY 3874 gagagaaatattaaagagaaatgaacaatagagatttttgggtttttaataaattat 3933

|||||
Db 210309 GAGAGAAATATTAAACAGAGAAATGAGACAATA-----TTTTGAGGTTTAAATTCG 210260
QY 3934 tgttaataataacatatagaataacttttttataaaataaacatgcacaacaataacacta 3993
Db 210259 TGGTAATGATATAACATATAAGAAACTTTTATTGAGATAATC--ACCAACAGTAACACTA 210202
QY 3994 tccgtctatctgcacagtttttcccccagggaagtgttttgccttttctcttttttttt 4053
Db 210201 TCATCTGTTCTGACACAGTCTCCCCCAGGAAAGTGTCTGTGCCAGTCTCTTCTT 210145
QY 4054 ttttttcatotttttgtttctctctcttttttccatcccttttttaatttttttaacag 4113
Db 210146 -----TCATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 210101
QY 4114 caatggagggaagttaacaatttttaa-tggaaagacatgtagagacaacaataacata 4172
Db 210100 CAATGGAGGAAGTTAAACAGTATTAAACAGGAAAGATGTCTCAGACGACGAATAATG 210041
QY 4173 agcaagactgcagcagcattataataatttttcaggggttttgcagctgaacataattcat 4232
Db 210040 AGCAAGATTGAGCAG-----TGTAACCATAGAGCTTTGAGGCTGATCATAGCTTGT 209988
QY 4233 tatccctcaaaaagttaccaccatcatcagaaaaaaa 4269
Db 209987 CATGCCCTCAAAATGCCTTCATACACACAGAAAAATAAAA 209951

RESULT 8
AB023157
LOCUS
DEFINITION Homo sapiens mRNA for KIAA0940 protein, complete cds.
ACCESSION AB023157
VERSION AB023157.1 GI:4589523
KEYWORDS
SOURCE
Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:hh04894.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.
XIII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
DNA Res. 6 (1), 63-70 (1999)
99246063
MEDLINE
2 (bases 1 to 5715)
Ohara,O., Nagase,T. and Kikuno,R.
Direct Submission
Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
FEATURES
Location/Qualifiers
1. .5715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hh04894"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
90. .2144
/gene="KIAA0940"
90. .2144
/gene="KIAA0940"
/codon_start=1
/product="KIAA0940 protein"
/protein_id="BAA76784.1"
/db_xref="GI:4589524"
/translation="MQDLLMDKSKTPQPOQOORQOQOPESSVSEAPSTPLSSET"
gene
CDS
```

```
PKPEENSAVPSALPAPAPNGPDKMOWESPILLGLSFHPQPPQPPPPQPPAPACAS
LSPFSFGTSTGTTNAVEDSFQGLTPVNGTLMFONFHHVNPVFGGTFSPQIGLAQT
QHQQQPPAPAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQ
QKSSSAVAAAAAASASSSWNTHSYNAASAPSPQWGLQAGRAAGVGV
GVGGVPSPLNPLSLPKKPPFSNNVIAAPKPPRAAPLTSKSWMEDNAFTDNGNLLPF
ODRSRYDFNLHSLNSLMDIRTDHEPLKGRMFINFHPGTDINMALNSRSLPFP
EDAFLLDDSHGQALSSGLSPYRCQNGVERVSKRVFVGGPLPDIDEIDETASPRFR
GLIUVDPHKAESKSYFPPKGVAFLLFOESSVOALIDACLEEDKLKLYLCVSPPIK
KPQIIPWNLSDSFVMDGSQLDPKRTIFVGVPRPLRAVELAMIMDKLYGGVCTAG
IDIDPELKPAGRAVAFNQOYSIAAISARFVQLOHNDIDKRVKPKYVLDQMCDE
CQGTGKGFAPFCANVTCLQOYCYCWASHTSRAGREFHRLPLVKEGGDRHRVFFR
WS"
BASE COUNT      1571 a  1201 c  1211 g  1732 t
ORIGIN

Query Match      10.1%; Score 430; DB 9; Length 5715;
Best Local Similarity 77.0%; Pred. No. 2.7e-82;
Matches 539; Conservative 0; Mismatches 155; Indels 6; Gaps 1;

QY 33 gtcagttcactgtttccgatgaaagattcttggaatgagccgtgggagtcagc 92
DB 1243 GTCGATCTTCTCTTTCCCTTTGAAGATGCCCTCTCGGATGATAGCCATGGTGAACAAG 1302
QY 93 ctcttcattagtgccgtgggttcacctcaactgcttcagtcaccagaaatgggagagagatgg 152
DB 1303 CTTTGTCTATCGGCTTAAGTTCCTCCACTCGCTGT-----CAAAATGGGGAACGAGTAG 1356
QY 153 aacgatattctcgaaggtttttagtggcggattgctccagacattgatgaagatgaga 212
DB 1357 AACGCTACTCTAGAAGGTGTTTGTGGAGGACTTCTCTCGATATTGATGAAGATGAGA 1416
QY 213 tcacagtagtttctgcgtgttgccctcgtgattgagttgagttgagtcataaagctgaga 272
DB 1417 TCACTGCGAGCTTTCGCGAGGTTTGGAGCTCTCTGCTAGTAGTGGCCCTCACAAGCTGAAA 1476
QY 273 gcaaatctattttccctcctaaaggtcattgcattcctgctgttccaagatgaaagctctg 332
DB 1477 GCAAGTCTATTTCCTTCCCTTAAGGCTATGCCCTTCTGCTGTTCCAGAGGAAGCTCAG 1536
QY 333 tgcaggtctcattgctgcattggaagatggaagatggaataactctaccctttgtgtatcaa 392
DB 1537 TACAAGCTTTGATAGATGCGCTGCGCTAGAGAAGATGGGAAACTCTACCTGTGTGTCTCAA 1596
QY 393 gtccactatcaaggataaagcagtcagattcgcccttggaatctcagtgacagtgact 452
DB 1597 GCCCCACATCAAGACACAGCAGTGCAAAATTCGACATGGAACCTGAAGTGACAGTGACT 1656
QY 453 ttgtgattggtgttcacagccacttgacccacgaaatactattttgttgggtgtgttc 512
DB 1657 TTGTAATGGATGTTCTTCAGCCTTTGGACCCACAGAAAACTATCTTTGTGGGGAGTTC 1716
QY 513 ctgacattacagagctgtgagacttgcgattgtaagtgaatcggtatcggtatcgagggtgtgt 572
DB 1717 CACGACCCCTTCGAGCTGTGTAATGCGAATGATCATGAGCCGCTTTGTACGGTGTGTCT 1776
QY 573 gctacgtggtattgataccgacctgagctaaaataaccacaaagagctggagagagtg 632
DB 1777 GCTATGCTGCATGTATACGACCCAGAGCTGAAGTACCCCAAGGTGCTGCGCCGCTGG 1836
QY 633 cgttcttaatacacagaggttacatagctgtatcagtcgcccgtttgttgcagctcagc 692
DB 1837 CATTCCTCAATCAGCAGAGTTACATTCGACGCCATCAGCGCTCGTTTGTGCAGCTTCAGC 1896
QY 693 atggagagatagataaacgggttaagccttatactacattt 732
DB 1897 ACAATGACATTGCAACACGGGTTTGAAGTAAGTAAAGCCATATGT 1936

RESULT      9
AC100917      71171 bp  DNA  linear  HTG 23-NOV-2001
LOCUS      Mus musculus clone RP23-71G18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
```

AC100917  
AC100917.1 GI:17059691  
HTG: HTGS\_PHASE0.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 71171)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-71G18  
Unpublished  
2 (bases 1 to 71171)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukigalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Hearford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,  
Jones,C., Kanat,A., Karatas,A., Kellis,C., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,W., Travis,N., Trigglio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zemбек,L., Zimmer,A. and Zody,N.

TITLE  
JOURNAL  
COMMENT

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L14574  
Center clone name: 71\_G\_18

-----  
\* NOTE: This record contains 87 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 718: contig of 718 bp in length  
719 818: gap of 100 bp  
819 1541: contig of 723 bp in length  
1542 1641: gap of 100 bp  
1642 2370: contig of 729 bp in length  
2371 2470: gap of 100 bp  
2471 3209: contig of 739 bp in length  
3210 3309: gap of 100 bp  
3310 4026: contig of 717 bp in length  
4027 4126: gap of 100 bp  
4127 4836: contig of 710 bp in length  
4837 4936: gap of 100 bp  
4937 5664: contig of 728 bp in length





Db 2077 GTGGGGGGTCCACGCTCTCTGAAGGCTTCGAACTGGCAATGATCATGATGATG 2136  
QY 560 taaggaggtgtgtaoagctggattgataccgacctgagctaaataaccacaaagga 619  
Db 2137 TAGCGTGGAGTATGCTATGCTGGAATGACACCGCATCGGAATTAAGATATCCAAAGGC 2196  
QY 620 gctggagaggttcgcttctctaatacaacaggttacatagctgctatcagtgcccgcttt 679  
Db 2197 GCTGGACGTGTGGCCCTCTCGAATCAGCAGACTACATAGCGGCCATCTCAGCCAGATT 2256  
QY 680 gttcagctgcagcatgagagatagataaacgggt 714  
Db 2257 GTCCAGCTGCAGCATGCCGATATAGACAGCGGT 2291  
RESULT 11  
AC014803/c 103054 bp DNA linear HTG 16-NOV-1999  
LOCUS Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
DEFINITION  
AC014803  
AC014803.1 GI:6436532  
VERSION HTG: HTGS\_PHASE2.  
KEYWORDS  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 103054)  
Direct Submission  
Adams, M. and Venter, J.C.  
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
This sequence was identified as CDM:10211121 by the submitter.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source  
1. .103054  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 29118 a 22259 c 22349 g 29328 t  
ORIGIN  
Query Match 7.1%; Score 304.6; DB 2; Length 103054;  
Best Local Similarity 70.6%; Pred. No. 2.5e-55;  
Matches 406; Conservative 0; Mismatches 169; Indels 0; Gaps 0;  
QY 140 gggagagatggaacgatattctgaaggtgttgtagcgattgacctccagacatt 199  
Db 45716 GGCGATGGCAGCGTCCCTCTCCCGAAGGTATGCTGTGGCGGTGCCACCGGACATC 45657  
QY 200 gatgaagatgagatcacagctagtttgcgtcgtttggccctctgatttggttgacct 259  
Db 45656 GATCAGGATGAGATTACCACCTTCGTTTCGGCGCTTCGGGCCATTGGTCGTCGATGGCCA 45597  
QY 260 cataaagctgagacgaactctatttccctccctaaagctatcattctcgtgtttcaa 319  
Db 45596 CACAAAGCGGAATCCAACTCGTATTTTCGGCCCAAGGGATGCTCTTCGCTGTCCAG 45537  
QY 320 gatgaagctctgtagcgtctctattgatgcattgacgattgaagaagatggaacctctac 379  
Db 45536 CACGAGACCATGTGCGACAGCTAATTGACTCGTGTCATCAGGATGAGGACAAAGCTGAT 45477  
QY 380 ctttgtgtatcaagtcacactatacaagataaagccagctccagattcgcccttggaatctc 439  
Db 45476 CTATGCGTTTCTTCGCCGACGATCAAGGATAAAGCAGTGCAGATTCGCTTGGCGCTG 45417  
QY 440 agtgacagtgactttgtgatgattggttcacagccacttgacccacgaaaaactattt 499

Db 45416 GCCGATCGGACTATGTGCTTGATGCTACCATGTCTACTGGACCCAGCAAAACGGTCTTT 45357  
QY 500 gttagtggtgttctctcgaaccattacgagctgttgagcttgcgattggttaataatggatcggcta 559  
Db 45356 GTGGCGCGGTGCCACGCTCTCTGAAGGCTTCGAACTGCGCAATGATCATGATGATGATG 45297  
QY 560 taaggaggtgtgtgtctacgctgggattgataccgacctgagctaaataaccacaaagga 619  
Db 45296 TACGGTGGAGTATGCTATGCTGGAATTGACACCGATTCGGAATTAAGATATCCAAAGGC 45237  
QY 620 gctggagaggttcgcttctctaatacaacaggttacatagctgctatcagtgcccgcttt 679  
Db 45236 GCTGGACGTGTGGCCCTCTCGAATCAGCAGACTACATAGCGGCCATCTCAGCCAGATT 45177  
QY 680 gttcagctgcagcatgagagatagataaacgggt 714  
Db 45176 GTCCAGCTGCAGCATGCCGATATAGACAGCGGT 45142  
RESULT 12  
AC010114/c 181720 bp DNA linear INV 03-FEB-2002  
LOCUS Drosophila melanogaster 3L BAC Rpci98-10P9 (Roswell Park Cancer  
DEFINITION Institute Drosophila BAC Library) complete sequence.  
AC010114  
AC010114.5 GI:15383757  
VERSION HTG.  
KEYWORDS  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 181720)  
REFERENCE  
AUTHORS  
Muzny, D., Scherer, S., Adams, M.D., Holt, R.A., Homs, C.A.,  
Gocayne, J.D., Tabor, P., Williamson, A., Homs, F.H.,  
Dugan-Rocha, S.D., Sodergren, E.S., Hodgson, A.H., Chen, R.C.,  
Ayala, M.A., Scott, G.S., Worley, K.W., Anammatides, P.G., Brandon, R.C.,  
Busam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K.,  
Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E.,  
Draper, H., Emery-Cohen, A., Ferriera, S., Garg, N.D.S., Houck, J.,  
Hostin, D., Howland, T.J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C.,  
Liu, W., Mattei, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B.,  
Nelson, K.A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V.,  
Schaefer, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M.,  
Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O.,  
Wheeler, D., Weinstock, G., Gibbs, R. and Venter, J.C.  
Direct Submission  
Unpublished  
2 (bases 1 to 181720)  
Worley, K.C.  
Direct Submission  
Submitted (13-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 181720)  
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Albrooks, S.L., Amarutunge, H.C., Are, J.R., Banks, T., Barbara, J.,  
Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Briefe, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,  
Burich, P., Burckett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Deigado, O.,  
Denn, A.L., Ding, Y., Dinh, H.D., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Farnhart, C., Edgar, D., Edwards, J.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,







gene

```
<10319..>29827
/gene="BcDNA:LD29930"
/notes="CG4999"
/map="66E3-66E4"
/db_xref="FLYBASE:FBan0004999"
/db_xref="FLYBASE:FBgn0029505"
join(23890..23955,24459..24557,25779..25907,28143..28502,
28563..28712)
/notes="BcDNA:LD29930"
/notes="BcDNA:LD29930 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004999"
/db_xref="FLYBASE:FBgn0029505"
/protein_id="AAF50354.1"
/db_xref="GI:7295026"
/translation="MVFDGVCWCAKYLICIFNFIFVLTGIIIFVGLWLVADKHSLLIA
LVLSERIEQTPQALAEQLAYLVLVIGAVFMFFMSFLGILGAMRESKRLISTYGTG
LIILLIAVAGISGAFKQKRAESKNFLOTITISYSLGENVDATSLMWNGLNFG
CCGINIDYHFDASPAWNGKNGNRTIPDACCIILDKVAKLVPRDEDCITNPDSNSFYKK
GCYEVFTWELIRORELVIIVAIVGIVLHVLIILIAFALCKAFKYNDRMRT"
complement(join(<30826..31252,31329..31903,32168..32359,
32429..32701,32759..32966,33028..34517,34833..34850))
/genes="CG5741"
/product="CT18050"
/db_xref="FLYBASE:FBan0005741"
/db_xref="FLYBASE:FBgn0035937"
complement(<30826..34850)
/genes="CG5741"
/map="66E3-66E4"
/db_xref="FLYBASE:FBan0005741"
/db_xref="FLYBASE:FBgn0035937"
complement(join(30826..31252,31329..31903,32168..32359,
32429..32701,32759..32966,33028..34517,34833..34850))
/genes="CG5741"
/notes="CG5741 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005741"
/db_xref="FLYBASE:FBgn0035937"
/protein_id="AAF50353.1"
/db_xref="GI:7295025"
/translation="MOVVKQVYHRRPQVSVPHCRRPVQVVEEAAPWAGVAAAGDSN
SSGTVRLIKGVKTQITIEIKSKSLPTAAIPTIIVADPPSTGWSNDPDSDTCLTIVNT
SISEPLDLDEHONGSDSVEAELLQHSISLGSQDVPTVAATGHSRSLRGLAHI
RSVKAKQAAPAMNKDVSQSGTVPATSVRTGFLRRNVVEAGSEPTSQOAEAI
ASGVKKRGILLARKDVEIESGVEVLEDMIPSTAVIHDSKDVANKRAEVDAGADA
DSRASLTPVAPGFTSLASIVEKSGVRDTRSRISFFSNLSMVSQWREKHQGAIA
FALPVLVLIILLMLPVPDFLRFVGFATLFFTVMDCCGHLRALLEDFFMLTTPER
VAFAPINPKMPCICEIPAVEBEKTIKYAGWMEINSYDPNTFSFLTRAVTVRLDGS
ILKMSGTNARIPKRWNNPEPIDRHKIPFTDHSYDLRDCRIELPLGLANKRFFNRK
YPTQLIISNRDVPDQSSVESOSEATVKTDAKEAKAASSGTTTPGGEELVDPAG
TVNQADLQOLRNVTNPKDELNDITVPCGDEVRLLLFSRCDREKEDWYRFTNASRGLV
HEQDQLQPMARFVSDQLQAAARQALLIVTGMKGAGVSEDEISAGKNDDETDOQTP
DGQFETIEEDLDTPKESFGDMINSANVARNSPDYVRFMAVYQSRRAKRODELEWK
GIDQSLFLGSPSGSIWVANLLGRCLWSLHDAMLHEKLEKLVQKLNSIKLPSFMEEV
IITNIVGLSPGLCHRVSPMLDERGVWIDADVTYEGLAHITVTTKLNLRLRSKTKS
PMATDNVPTTSMLEPOSQDLRSLSEVNSENVLYDSDAESGGSTSESPSPPCNAPEN
AACTEFQNSPGNARLFIKIVDIRAASNLFOYATELPYVQRAMENNNANITLRVLDKG
LVARGHILNPPSPSRVWSFGPPRLNISTKFPQVGDKSVDSIVTNVIESLKCEAVN
KFLVYPMNVDFSIFLUSNPNFDDQNAAN"
complement(join(39649..42284,45592..45653,46647..47044,
47421..47523,50583..50980))
/genes="CG5735"
/product="CT17928"
/db_xref="FLYBASE:FBan0005735"
/db_xref="FLYBASE:FBgn0035938"
complement(join(39649..42284,45592..45653,46647..47044,
47421..47523,51697..52020))
/genes="CG5735"
/product="CT18021"
/db_xref="FLYBASE:FBan0005735"
/db_xref="FLYBASE:FBgn0035938"
complement(join(39649..42284,45582..46293))
/genes="CG5735"
```

mRNA

mRNA

mRNA

Query Match

```
Best Local Similarity 7.1%; Score 304.6; DB 3; Length 287018;
Matches 406; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 140 ggggagagagtgaaacgatatctcgaagggtgttttagcgcgattgctccagacatt 199
DB 41455 GCGGATGGCAGGCTCGCTTCCTCCGGAAGGTAATTCGTGGCGGCTGCCACCGACATC 41396
QY 200 gatgaagatgagatcacagctagttttcctcgcttttggccctctgattgtggatggcct 259
DB 41395 GATGAGGATGAGATTACCACCTTCGTTTCGGCGCTTCGGGCCATTGTCGTCGATTGGCCA 41336
QY 260 cataaagctgagagcaaatcctatttctctctctctctctctctctctctctctctctctct 319
DB 41335 CACAAGCGGAAATCCAAAGTCGTAATTTCCGCCCAAGGGATATGCCCTCCTGCTGTTCAG 41276
QY 320 gatgaagctctgtgcagctctcattgatgatgatgatgatgaagaagatgaaactctac 379
DB 41275 GACGAGAGCAGTGTGCAGCAGCTAATTGACTCGTCATCCAGGATGAGGACAAGCTGTAT 41216
QY 380 ctttgtgtatcaagtcctccactcaaggataagcagctccagcttcgcttggactctc 439
DB 41215 CTATGCGTTTCTTCGCGCAGCATCAAGGATAAGGCGAGTCGAGATTCGTCCTTGGCGCTG 41156
QY 440 agtgacagtgaactttgtgatgatggttcacagcgaacttgacccacgaaactatatt 499
DB 41155 GCGGATGCGGAGTATGTGTTGATGTACCATGTCACTGCAGCCACGCAAAACGGTGT 41096
QY 500 gttgggtgttctcgcagcattagagctgtggagcttgcgatgtaagtgcgcta 559
DB 41095 GTGGCGCGGTGCGCCTCTCTGAAGCCTTCGAAGCTTGCAGTGCAGTATGATGATGATG 41036
QY 560 tacggaggtgtgtacgctgggattgatcacgcacctgagctctgctctgctctgctctgctct 619
DB 41035 TAGGGTGGAGTAGTGTATGCTGGATTGACACCGATCCGGAATTAAGTATCCAAAGGC 40976
QY 620 gctgggagagtgctgtctctctctctctctctctctctctctctctctctctctctctctct 679
DB 40975 GTGGAGCTGTGGCTTCTCGAATCAGCAGAGTACATAGCGGCCCTCTCAGCCAGATT 40916
QY 680 gttcactcagcatgagagatagataaacgggt 714
DB 40915 GTGCAGCTGAGCATGGCGATATAGACAAGCGGCT 40881
RESULT 14
AC009968/c 205307 bp DNA linear HTG 26-JUL-2000
LOCUS Homo sapiens chromosome 19 clone RP11-402N14, WORKING DRAFT
DEFINITION SEQUENCE, 22 unordered pieces.
ACCESSION AC009968
VERSION AC009968.6 GI:9454634
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 205307)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 205307)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jul 26, 2000 this sequence version replaced gi:8748919.
----- Genome Center -----
```



```
QY 1888 tcaactcaactctgctcccggttcaagtgatctcctgcttagctcccgagtagg 1947
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122959 TCACTGCAACTCTGCTCCAGGTTCAAGCAATTCCTGCTCAGTCACCTGAGTAGC 122900
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1948 tgaattacaggcatataccaccatcccaacaacatgtttgtattttttagtgagacagg 2007
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122899 TGGGAATACAGCGCCACCACCAATGCCAGCAAAATTTTGTATTTTGTAGTAGACAGG 122840
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2008 gtttacccgtgtgtggtcagccaggttcaaaactcctgactcaaggagatccacctgcctc 2067
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122839 GTTTCACCATGTTGCCAGGCTGTTTCGAACCTCTGAGCTCAGGTGATCGGCCACCTTC 122780
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2068 accccctcaagtgctgggattacagcagcaccacaccatgcctggcctacttggt 2127
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122779 GGCTCCCAAGTGTCTGGGATTACAGATGAGGCCACCTCGCTGCCCAAAAGCTCTAA 122720
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2128 ttatttgacacataaaataactcatctcactgccttatttccaaacataagtttcaga 2187
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122719 TTTTATG-----AGAAACTCTGAGGACAGCAATCTAGTCAATTGTTAATGAATNN 122669
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2188 gctggtggattggtcattagaaattcagactgaattgttctcctcgaatgaatactc 2247
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122668 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122609
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2248 ttgccagtggtcactctgtagacattatgagcagcctagagccagcagaagccca 2307
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122608 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122554
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2308 gtctctccttgcctgtctctcctggttcgtgacactcttctctctcttcttcttct 2367
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122553 TTCCTTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 122494
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2368 ttatttttttagtaaaaaatttttttagaggagggtctcaactctgcaacccaggt 2427
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122493 TTTCTCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 122434
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2428 ggaacacagaatcacatcatgactcactgctgtctctctctctctctctctctctc 2483
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122433 GGACTCGATGGTCAATCTCGGCTCACTGCAAGCTCCAGCTCCAGGTTTCAGGCC 122378
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 15
AC004973/c
LOCUS AC004973 97037 bp DNA linear PRI 21-DEC-1999
DEFINITION Homo sapiens PAC clone RP5-113911 from Xq23, complete sequence.
ACCESSION AC004973
VERSION AC004973.1 GI:3694660
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 97037)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 97037)
AUTHORS Courtney,L., Langston,Y. and Drone,K.
TITLE The sequence of Homo sapiens PAC clone RP5-113911
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 97037)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 97037)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 97037)
```

AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Waterston,R.  
Direct Submission  
Submitted (18-MAR-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 97037)  
Waterston,R.  
Direct Submission  
Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 3, 1998 this sequence version replaced gi:3213024.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_DJ1139101  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

#### MAPPING INFORMATION:

This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the chromosome X mapping group  
at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK.  
Further information can be found at  
<http://www.sanger.ac.uk/HGP/ChrX/>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by  
Pietter de Jong and coworkers at the Roswell Park Cancer Institute  
(<http://bacpac.med.buffalo.edu>) using the method described by  
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from  
one male donor.  
The clone may be obtained either from Genome Systems, Inc.  
(<http://www.genomesystems.com>) or Research Genetics, Inc.  
(<http://www.resgen.com>); or from Pieter de Jong.  
VECTOR: pCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-555N2; the clone sequenced  
to the right is RP3-404F18, 200 bp overlap. Actual start of this  
clone is at base position 1 of RP5-113911; actual end is at 13269  
of RP3-404F18.

FEATURES	Location/Qualifiers
source	1..97037
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="Xq23"
	/clone="RP5-113911"
	/clone_lib="RPCI-5"
	530..801
repeat_region	/rpt_family="L2"
	830..1035
repeat_region	/rpt_family="Alu"
	2179..2349
repeat_region	/rpt_family="MER1_type?"
	3434..3489
repeat_region	/rpt_family="MIR"
	3526..3739
repeat_region	/rpt_family="MIR"
	4681..4852

repeat_region	/rpt_family="MIR"	5395. .5961	20738. .20828	repeat_region	/rpt_family="Achobo"	20848. .21707
repeat_region	/rpt_family="MER4-group"	6293. .6393	21717. .22111	repeat_region	/rpt_family="L2"	22163. .22378
repeat_region	/rpt_family="L1"	6394. .7105	22163. .22378	repeat_region	/rpt_family="L2"	22866. .23031
repeat_region	/rpt_family="L1"	7104. .7211	22866. .23031	repeat_region	/rpt_family="L1"	23060. .23339
repeat_region	/rpt_family="L1"	7342. .7520	23060. .23339	repeat_region	/rpt_family="Alu"	23409. .24201
repeat_region	/rpt_family="L1"	7542. .7644	23409. .24201	repeat_region	/rpt_family="L2"	24203. .24554
repeat_region	/rpt_family="Alu"	7645. .8674	24203. .24554	repeat_region	/rpt_family="Achobo"	24572. .25105
repeat_region	/rpt_family="Retroviral"	8682. .9314	24572. .25105	repeat_region	/rpt_family="Achobo"	25106. .25479
repeat_region	/rpt_family="Retroviral"	9315. .9368	25106. .25479	repeat_region	/rpt_family="L1"	25480. .25768
repeat_region	/rpt_family="7SLRNA"	9371. .10157	25480. .25768	repeat_region	/rpt_family="Alu"	25769. .25985
repeat_region	/rpt_family="Retroviral"	10165. .10817	25769. .25985	repeat_region	/rpt_family="L1"	
repeat_region	/rpt_family="Retroviral"	10818. .10973				
repeat_region	/rpt_family="Alu"	10974. .11111				
repeat_region	/rpt_family="L1"	11238. .11612				
repeat_region	/rpt_family="MaLR"	11972. .12283				
repeat_region	/rpt_family="Alu"	12570. .12854				
repeat_region	/rpt_family="L1"	12859. .13433				
repeat_region	/rpt_family="MER1_type"	13435. .13564				
repeat_region	/rpt_family="Alu"	13580. .13845				
repeat_region	/rpt_family="MER1_type"	13850. .14118				
repeat_region	/rpt_family="L1"	14119. .14411				
repeat_region	/rpt_family="Alu"	14412. .14813				
repeat_region	/rpt_family="L1"	14801. .15447				
repeat_region	/rpt_family="L1"	16501. .16654				
repeat_region	/rpt_family="L2"	17412. .17533				
repeat_region	/rpt_family="MER1_type"	17534. .17793				
repeat_region	/rpt_family="Alu"	17794. .17912				
repeat_region	/rpt_family="MER1_type"	17933. .18249				
repeat_region	/rpt_family="L1"	18293. .18358				
repeat_region	/rpt_family="MIR"	18348. .18476				
repeat_region	/rpt_family="L2"	18483. .18908				
repeat_region	/rpt_family="MaLR"	18932. .19227				
repeat_region	/rpt_family="Alu"	19390. .19604				
repeat_region	/rpt_family="L2"	19606. .20227				
repeat_region	/rpt_family="Achobo"	20228. .20477				
repeat_region	/rpt_family="Alu"	20478. .20721				
repeat_region	/rpt_family="Achobo"					

Query Match

5.6%; Score 238; DB 9; Length 97037;

Best Local Similarity 63.6%; Pred. No. 5.1e-41;

Matches 412; Conservative 0; Mismatches 230; Indels 6; Gaps 3

QY	1823	tttgtttgttgagatggagttctctccgtcacccaggtcggagtgagtgacagtc	1882
DB	54938	TTTGTCTTTTTCAGACAGAGTCTTGCTGTCATCCAAAGCTGGAGTGCAGTGCACGATC	54879
QY	1883	tcagctcaactcaaacctctgctccgggttcaagtattctctgcttagctcccgga	1942
DB	54878	TCAGCTCACTCAACCTCTGCTCTGCTGTTCAAGCAATCTCTGCTCAGCCTCCCGA	54819
QY	1943	gtagtggaattacagggatataccaccatgcccaacaaatgtttgtatttttagtgag	2002
DB	54818	GTAGCTGGGACTACAGGGCGCATGCCACCACGCCAGCTAAATTTGTATTTTAGTAGAG	54759
QY	2003	acaggggtttcaacgtgtgttggttcagggcagtttcaaacctctcgaactcaaggatccac	2062
DB	54758	ATGGGGTTTTCACCATATTGGTCAGGCTGGTCTTGAACCTCCTGAGCTCAGATGATCCACCT	54699
QY	2063	gcctcacccctcaaaagtcgtgggttacagggcatgagccacc-caccatgcctggccta	2121
DB	54698	GCTCGGCTCCCAAAGTCTGGGATTTACAGCGGTGAGCCACCGCGCGGCTCTTCTA	54639
QY	2122	cttggtttttatgcacactaaaaatacctacatctcactgccttattccacataagt	2181
DB	54638	GCTATTTTTGGAAATGTACAATTTGATTACGTTTAACCTTTAATCACCTACTGATGATTTGAAA	54579
QY	2182	ttcagagctgtgggattgggtcattagaattcagacactgaattgttctctcctaata	2241
DB	54578	ACCAGGCTCTTATTTCTTAAAGTGTGATTTGTGTCATTTTATATACTGTTTCTTATTT	54519
QY	2242	aatcctttgcccagtggttcactgttagacatttagagcagcagtagagccaga	2301
DB	54518	ATGCTCTGTGTTCTTTGTTTTCCTATTGCTTTTCTGCTTCCCTGGTTTTAAATTTG	54459
QY	2302	agc---ccagtgctcctctatgcctgctctctctctctctctctctctctctctctct	2356
DB	54458	AGCTTTTTCATATGATTCCATTTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	54399
QY	2357	cttttgctacttt	2416
DB	54398	TATACTTTTATATATATACATTTTAAATTAATTTATTTTATTTTATTTTATTTTATTTT	54339
QY	2417	tcacccaggttgagcagaatcacatcatgactgaactgcatgttc	2464
DB	54338	CCACCCAGGCTGGAGTGGCTCAATCTCGGCTCACTCGGCTCACTCGACGCTC	54291

Search completed: September 1, 2002, 09:45:29  
Job time: 14874 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2002, 06:53:09 ; search time 469.45 seconds  
(without alignments)  
15638.569 Million cell updates/sec

Title: US-09-730-559b-1

Perfect score: 4276

Sequence: 1 ttctacggtttttccctgc.....atcagaaaaa 4276

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4276	100.0	4276	21	Human IgA nephropathy
2	544	12.7	544	22	Human foetal liver
3	544	12.7	544	22	Probe #11103 used
4	356.4	8.3	1290	23	DNA encoding novel
5	304.6	7.1	2748	23	Drosophila melanog
6	304.6	7.1	3523	23	Drosophila melanog
7	304.6	7.1	3597	23	Drosophila melanog
8	304.6	7.1	5045	23	Drosophila melanog
9	304.6	7.1	13332	23	Drosophila melanog

10	304.6	7.1	14372	23	ABL09878
C 11	299	7.0	305	19	AAV49889
C 12	299	7.0	305	21	AAV49889
C 13	225.2	5.3	11003	24	AAV49889
C 14	225.2	5.3	11003	24	AAV49889
C 15	225	5.3	225	22	AAV49889
C 16	225	5.3	225	22	AAV49889
C 17	223.8	5.2	6418	22	AAV49889
C 18	223	5.2	32221	22	AAV49889
C 19	223	5.2	32221	22	AAV49889
C 20	222.4	5.2	4374	22	AAV49889
C 21	221.8	5.2	31168	22	AAV49889
C 22	221.8	5.2	31168	22	AAV49889
C 23	221.8	5.2	31168	22	AAV49889
C 24	221.8	5.2	31168	22	AAV49889
C 25	221.8	5.2	31168	22	AAV49889
C 26	221.8	5.2	31168	22	AAV49889
C 27	221.8	5.2	31168	22	AAV49889
C 28	221.8	5.2	31168	22	AAV49889
C 29	221.8	5.2	31168	22	AAV49889
C 30	221.8	5.2	31168	22	AAV49889
C 31	221.8	5.2	31168	22	AAV49889
C 32	221.8	5.2	31168	22	AAV49889
C 33	221.8	5.2	31168	22	AAV49889
C 34	220.2	5.1	22465	22	AAV49889
C 35	219	5.1	27960	22	AAV49889
C 36	219	5.1	27960	22	AAV49889
C 37	218.8	5.1	825	22	AAV49889
C 38	218.6	5.1	9620	22	AAV49889
C 39	218.6	5.1	10483	22	AAV49889
C 40	218.4	5.1	27884	22	AAV49889
C 41	218	5.1	46107	22	AAV49889
C 42	217.8	5.1	8858	22	AAV49889
C 43	217.6	5.1	160552	22	AAV49889
C 44	217.2	5.1	7582	22	AAV49889
C 45	217.2	5.1	7582	22	AAV49889

#### ALIGNMENTS

RESULT 1

AAV49889

ID AAV49889 standard; cDNA; 4276 BP.

AC AAV49889

XX AAV49889

DT 10-NOV-2000 (first entry)

XX AAV49889

DE Human IgA nephropathy-associated cDNA INP303A.

XX AAV49889

XX AAV49889

KW IgA nephropathy-associated protein; diagnosis; treatment; antisense;

human; INP303A; ss.

XX AAV49889

OS Homo sapiens.

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX AAV49889

XX  
DR WPI; 2000-097328/08.  
DR P-PSDB; AAW90737.

DNA sequences preferentially expressed in IgA nephropathy patients, PT proteins encoded by them, and antibodies to those proteins

PS Claim 1; Page 76-83; 180pp; Japanese.

This invention describes novel DNA sequences preferentially expressed in IgA nephropathy patients, and DNA sequences stringently hybridizing to them. Independent claims cover diagnostic reagents for IgA nephropathy incorporating the antisense sequences; the treatment of IgA nephropathy using the antisense sequences for mRNA inhibition; proteins associated with IgA nephropathy, containing sequences encoded by the DNA sequences; antibodies recognizing these proteins; the production of the proteins by culture of host cells transformed with DNA encoding them; diagnostic reagents for IgA nephropathy containing the antibodies; and compositions for the treatment of IgA nephropathy which contain the antibodies. The products of the invention can be used for the diagnosis and treatment of IgA nephropathy. This sequence encodes the human IgA nephropathy-associated protein INF303A which is described in the method of the invention.

Sequence 4276 BP; 1149 A; 854 C; 854 G; 1419 T; 0 other;

```
Query Match      100.0%; Score 4276; DB 21; Length 4276;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 4276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 ttctaccggttttttccccctgctttcttatccagggtcagttcttccactgtttccgatggaaga 60

Db	601	gctaaatacccaaaaggagctgggagagtgctgcttctctaataccaacagagttacatagc	660
Qy	661	tgtataagtgcocgctttgttcaagctgcagcatggagagatagataaaacgggttaagcct	720
Db	661	tgtatacagtcocgctttgttcaagctgcagcatggagagatagataaaacgggttaagcct	720
Qy	721	tatactacatttggaaaaattctagaaattggttcctctaataatgtgtattaccaaatattag	780
Db	721	tatactacatttggaaaaattctagaaattggttcctctaataatgtgtattaccaaatattag	780
Qy	781	aacgggaggaattttatgacaataaagttgacagctgcacaaatttgcctcatagagttaatta	840
Db	781	aacgggaggaattttatgacaataaagttgacagctgcacaaatttgcctcatagagttaatta	840
Qy	841	tgtctataatcacatgaataatgtcctcatgaattcttttaattcttccagtttttttgagt	900
Db	841	tgtctataatcacatgaataatgtcctcatgaattcttttaattcttccagtttttttgagt	900
Qy	901	agcctaatacagaacactacaatttacttgagtttaatttaattcttcttaacttccattca	960
Db	901	agcctaatacagaacactacaatttacttgagtttaatttaattcttcttaacttccattca	960
Qy	961	atctcaatcatcgcctcaattcatcaacttagtttgttaagctcatcctaataaattttact	1020
Db	961	atctcaatcatcgcctcaattcatcaacttagtttgttaagctcatcctaataaattttact	1020
Qy	1021	gaatcctttgttctgtgtttatatcaagtatatacaaacagagaatgccttgaggtttctctgc	1080
Db	1021	gaatcctttgttctgtgtttatatcaagtatatacaaacagagaatgccttgaggtttctctgc	1080
Qy	1081	ccctttttttgtgttttttaactcctgggacatagggaagacctcagcaagcctctattt	1140
Db	1081	ccctttttttgtgttttttaactcctgggacatagggaagacctcagcaagcctctattt	1140
Qy	1141	ctcaaatgaatttactcacagattcttttttttttttttttttttttttttttttttttttt	1200
Db	1141	ctcaaatgaatttactcacagattcttttttttttttttttttttttttttttttttttttt	1200
Qy	1201	acctctcaacgatttatctcttagcttggtgtttcatgtattctcaacaaaagtttttagtgc	1260
Db	1201	acctctcaacgatttatctcttagcttggtgtttcatgtattctcaacaaaagtttttagtgc	1260
Qy	1261	ttaggccaagaagctcctgtctctatgagtttattctctcagcatagaaactgtatacact	1320
Db	1261	ttaggccaagaagctcctgtctctatgagtttattctctcagcatagaaactgtatacact	1320
Qy	1321	tgcagttactactcacagagtgtggcctgtggactgcacctccagctctgttaaacttagtttgt	1380
Db	1321	tgcagttactactcacagagtgtggcctgtggactgcacctccagctctgttaaacttagtttgt	1380
Qy	1381	agtgagataggaatttagaccagaatgtgtaatacacaacttactctgggcacaaatgtttg	1440
Db	1381	agtgagataggaatttagaccagaatgtgtaatacacaacttactctgggcacaaatgtttg	1440
Qy	1441	gtcagctgvcgatttttttttcatagaagaagcctttattgatgagggaagcaataatttg	1500
Db	1441	gtcagctgvcgatttttttttcatagaagaagcctttattgatgagggaagcaataatttg	1500
Qy	1501	atttatattttgggtgcacctttttattctatggcacactggcactttcatgcagtctga	1560
Db	1501	atttatattttgggtgcacctttttattctatggcacactggcactttcatgcagtctga	1560
Qy	1561	cttttgatatccactcactctgaggcatgtgctaaataatgattgttttatcgttgtttctc	1620
Db	1561	cttttgatatccactcactctgaggcatgtgctaaataatgattgttttatcgttgtttctc	1620
Qy	1621	tcaattcaagatgtaaaaatcatcaagtcagtcagtcagttttgtcttttatgtttcaatgt	1680
Db	1621	tcaattcaagatgtaaaaatcatcaagtcagtcagtcagttttgtcttttatgtttcaatgt	1680
Qy	1681	catgtacagcttacttcaactgcgcagtaaaaaaatttaagaatagtcggtgtctcatcctcaa	1740
Db	1681	catgtacagcttacttcaactgcgcagtaaaaaaatttaagaatagtcggtgtctcatcctcaa	1740



```
|||||
Db 3901 caatagagcatttgggttttaattaaattattgttaataataacataaagaatac 3960
Qy 3961 ttttataaaaaaacacatacaacatacactatcggtctctatctgacagttttccccc 4020
Db 3961 ttttataaaaaaacacatacaacatacactatcggtctctatctgacagttttccccc 4020
Qy 4021 gggaaagcgttttgccttttcccttttcttttttttttttttttttttttttttttct 4080
Db 4021 gggaaagcgttttgccttttcccttttcttttttttttttttttttttttttttttct 4080
Qy 4081 cttttttccatcccttttttaatttttttaacagcaatggaggaagttaacaattttta 4140
Db 4081 cttttttccatcccttttttaatttttttaacagcaatggaggaagttaacaattttta 4140
Qy 4141 ggaagagacatgttagagcaacaacaatgataagaagaactgagcagcattataatt 4200
Db 4141 ggaagagacatgttagagcaacaacaatgataagaagaactgagcagcattataatt 4200
Qy 4201 tttcagggttttgaggcgtgaacataatttcattatccctcaaaaagttaaccaccac 4260
Db 4201 tttcagggttttgaggcgtgaacataatttcattatccctcaaaaagttaaccaccac 4260
Qy 4261 gaaaaaaaaaaaaa 4276
Db 4261 gaaaaaaaaaaaaa 4276

RESULT 2
ABA62438
ID ABA62438 standard; DNA; 544 BP.
XX
AC ABA62438;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #10743.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
DR WPI; 2001-483447/52.
XX
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX
PS Claim 1; SEQ ID NO 10743; 639pp + sequence listing; English.
XX
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
```

```
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fip.wipo.int/pub/published_pot_sequences.
XX
SQ Sequence 544 BP; 128 A; 113 C; 125 G; 178 T; 0 other;

Query Match 12.7%; Score 544; DB 22; Length 544;
Best Local Similarity 100.0%; Pred. No. 3e-91;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3301 agcattgctggcctagtcactgaaaaatgtaaacctcttattttgattgcagtggaagt 3360
Db 1 agcattgctggcctagtcactgaaaaatgtaaacctcttattttgattgcagtggaagt 60
Qy 3361 taaccataatctcttgatgatcagctgtgtgatgaatgcaggggcccgctgtggggg 3420
Db 61 taaccataatctcttgatgatcagctgtgtgatgaatgcaggggcccgctgtggggg 120
Qy 3421 gaaattgtccatttttctgtgctaattgttacctgtctgcagtattactgtgaattg 3480
Db 121 gaaattgtccatttttctgtgctaattgttacctgtctgcagtattactgtgaattg 180
Qy 3481 ctgggctgctatccattctctgtgctgaggaattccacaagccctggtgaaggagg 3540
Db 181 ctgggctgctatccattctctgtgctgaggaattccacaagccctggtgaaggagg 240
Qy 3541 cggtagccgcctcgccatatttccattccgctggaactaaaggataaactgcagtgcct 3600
Db 241 cggtagccgcctcgccatatttccattccgctggaactaaaggataaactgcagtgcct 300
Qy 3601 tttcaggcctcagaataagtgcaactctctgttcaattctgaccccttcccaacctctc 3660
Db 301 tttcaggcctcagaataagtgcaactctctgttcaattctgaccccttcccaacctctc 360
Qy 3661 acgctggcatgctctttgttagcagctgtgaacttaactatagatataaagaagaatg 3720
Db 361 acgctggcatgctctttgttagcagctgtgaacttaactatagatataaagaagaatg 420
Qy 3721 cctataataataggctgtttgttagattctgtcactgcaacaataatgaactcctttt 3780
Db 421 cctataataataggctgtttgttagattctgtcactgcaacaataatgaactcctttt 480
Qy 3781 cgtattgccatcggttgatcgatgggaagttttattctctgtttgtctggaacaaaggga 3840
Db 481 cgtattgccatcggttgatcgatgggaagttttattctctgtttgtctggaacaaaggga 540
Qy 3841 tcca 3844
Db 541 tcca 544

RESULT 3
AAI42417
ID AAI42417 standard; DNA; 544 BP.
XX
AC AAI42417;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #11103 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
```

```
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 11103; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 544 BP; 128 A; 113 C; 125 G; 178 T; 0 other;
SQ

Query Match 12.7%; Score 544; DB 22; Length 544;
Best Local Similarity 100.0%; Pred. No. 3e-91;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3301 agcattgctggcctagctacactgaaaatgaaactcttattttgattgcaggtggaagt 3360
Db 1 agcattgctggcctagctacactgaaaatgaaactcttattttgattgcaggtggaagt 60

QY 3361 taagccatatgttgatgacagtgtgatgaatgcagggcccgctgtggagg 3420
Db 61 taagccatatgttgatgacagtgtgtgatgaatgcagggcccgctgtggagg 120

QY 3421 gaaattgtccattttctgtcctaattgtacctgtctgcagattactgtgaatttg 3480
Db 121 gaaattgtccattttctgtcctaattgtacctgtctgcagattactgtgaatttg 180

QY 3481 ctgggctgtatccattctcgtctggcagggaattccacaagccctcgtggaaggagg 3540
Db 181 ctgggctgtatccattctcgtctggcagggaattccacaagccctcgtggaaggagg 240

QY 3541 cggtgaccgcctcgcgcataatttcacccgtggaactaaaggataactgcagtgctcat 3600
Db 241 cggtgaccgcctcgcgcataatttcacccgtggaactaaaggataactgcagtgctcat 300

QY 3601 ttccaggcctcagaataagtgcaactctctgttcattctcgtgaccctctcctcaacctcttc 3660
Db 301 ttccaggcctcagaataagtgcaactctctgttcattctcgtgaccctctcctcaacctcttc 360

QY 3661 acgtgcatgctctttgtgacagctctgtaacttaactatagataatgaagaatga 3720
Db 361 acgtgcatgctctttgtgacagctctgtaacttaactatagataatgaagaatga 420

QY 3721 cctataataggtgtttgttagattctgtgcactgcgaacaaatagtaactcctttt 3780
Db 421 cctataataggtgtttgttagattctgtgcactgcgaacaaatagtaactcctttt 480

QY 3781 cgtattgccatcgggttgcatggaagttttattctcttctgttctgtggaaccaagagga 3840
Db 481 cgtattgccatcgggttgcatggaagttttattctcttctgttctgtggaaccaagagga 540

QY 3841 tcca 3844
Db 541 tcca 544
```

```
RESULT 4
AAS73462
ID AAS73462 standard; cDNA; 1290 BP.
XX
AC AAS73462;
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #9266.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG09275.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 9266; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1290 BP; 345 A; 273 C; 281 G; 391 T; 0 other;
SQ
```

```
Query Match 8.3%; Score 356.4; DB 23; Length 1290;
Best Local Similarity 72.9%; Pred. No. 1.2e-56;
Matches 529; Conservative 0; Mismatches 186; Indels 11; Gaps 5;

QY 33 gtcagcttcctcactgtttccgattggaagatggattcttgatgatgccgtgggagtcagc 92
Db 137 gtcggtcttcctcctcttccaatagatgatggcttgcttgatggtcacagtgtatca-- 194
QY 93 ctcttcattagtgccctgggttcacctcactcactgcttcagtcaccacagatggggagagtg 152
```

Db 195 -agttggaggttttaattcaaccacctgttattcagctcaccacaaatggagagcgaatag 253  
QY 153 aacgatatctcogaaggtgttttagcggtattgcctccacagacattgatgaagatgaga 212  
Db 254 aacgctctctcgaagaagttttgtgtgtgcttccctccagatattgatgaagatgaaa 313  
QY 213 tcacagctagtttctgcttggccctctgattggtgattggtgcctcataaaagctgaga 272  
Db 314 taacgcctagcttcagaagatttggcccttggtagtagattggcctcataaagcagaaa 373  
QY 273 gaaatctatttctctcctaagagctatgcattctctgcttttcaagatgaaagctctg 332  
Db 374 gcaagctctatttccacaaagagctatgcatttctcttcccaagaagagagctcag 433  
QY 333 tgcagctctcattgatgcattgattgaagaagatgg-aaacctctacctttgtgtatca 391  
Db 434 ttacggcactcattgatgctgttattgaagaagatggaaaaactctacctgtgtgttcc 493  
QY 392 agtccactat----caaggataagcagctccagattcgg--ccttggaaatctcaagtac 445  
Db 494 tagccctactatccaaggacaaacccagttccaaatacgtcccttggaaatttaagtgt 553  
QY 446 agtgaacttggatggatggttcacagccacttgaccacagaaaaactatatttgggt 505  
Db 554 agtgaatttgaatggatggttccagcccttggatcccccgaaaaaaatttttgggtga 613  
QY 506 ggtgtctctcagaccattacagctgtgagcttgcgattggtgtaattggatcggtatacga 565  
Db 614 ggtgtctcagccattaaagggtctgtaacctgtctatgatcattgagccgctgtatgt 673  
QY 566 ggtgtgtcactgctggattgatcacgaccctgagctaaataaccacaa-aaggagctgg 624  
Db 674 ggaatttggatgcagaatgtatcacagatcctgagctaaataaccacaaagctcctgg 733  
QY 625 gagagttgcgtctctcctaacaacagagttacatagctgctatcagtgcccgctttgtca 684  
Db 734 gcagtgctcttctcccaatcagcagagctatattgctgcccattagtgctggtttgtca 793  
QY 685 gctcagcagtgagagatagataaacgggttaagccttatactacattttggaataattcta 744  
Db 794 gcttcagcagtggtatattgataaacgttggaggttaagcccatatgtgctagatgacca 853  
QY 745 gaaatg 750  
Db 854 gatgtg 859

## RESULT 5

ABL09795

ID ABL09795 standard; cDNA; 2748 BP.

AC ABL09795;

XX

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23867.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN W0200171042-A2.

XX

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

XX

XX 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE ) PE CORP NY.

XX

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR P-PSDB; ABB65692.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

XX

PS Claim 1; SEQ ID NO 23867; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
sequences (ABL01840-ABL16175) and the encoded proteins  
(ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 2748 BP; 709 A; 674 C; 694 G; 671 T; 0 other;

## Query Match

Best Local Similarity 7.1%; Score 304.6; DB 23; Length 2748;

Matches 406; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 140 ggggagagagtggaacgatattctcgaaaggtgttttagcgaggttgcctccagacatt 199

Db 942 ggcgatggcacgctgcgtctctcccggaaggtattcgttggcgctgccaccgacatc 1001

QY 200 gatgaagatgagatcacagctagtttcgttcgttgcctctctgattgtagtgagct 259

Db 1002 gatggagtagattaccacttcgtttcggtcggtcggttcgttcgttcgttcgttc 1061

QY 260 cataaagctgagagcaaatcctatttctcctaaagctatgcattcctctgtttcaa 319

Db 1062 cacaagggcggaatacgaagtcgtatttccgcccgaaggatgccttctctgtctccag 1121

QY 320 gatgaagctctgtgcaggtctcattgatgcattgaagaagatggaacactctac 379

Db 1122 gacgagagcagtgtagcagcgttaattgactcgtgcacacgtaggagacagctgtat 1181

QY 380 ctttgtgtatcaagtcaccactatcaagataaagcagtcacagattcgccttggaatctc 439

Db 1182 ctatgcgttcttcgcagcatcaagataaaggcagtgacagattcgtcttgcgcctg 1241

QY 440 agtgacagtgcatttgcaggtgttcacagccacttgaccacacgaaactatatt 499

Db 1242 gccgatgcgactatgtgcttgatgctaccatgtcacttgacccacgcaaacggtgtt 1301

QY 500 gttgtgtgttctctcgaccattacagagctgtgagcttgcgattgtagtgagctgcta 559

Db 1302 gtggcggtgcgtgccacgtctctgaagccttcgaactggaatgatcatgatgatg 1361

QY 560 tacggaggtgtgctacgctgggattgataccagccctgagctaaataaccacaaagga 619

Db 1362 tacggtggagtgatgctatgctggaaattgaccccatccggaattaaagttatccaa 1421

QY 620 gctgggagagttgcgttctcctaacaacagagattacatagctgtatcagtgccgcgttt 679

Db 1422 gctggacgtgtggccttctcgaatcagcagagctacatagcggccatctcagcagatt 1481

QY 680 gttcagctgcgacgtgagagatagataaacgggt 714

Db 1482 gtgcagctgcagatggcgatatagacaaagcgggt 1516

RESULT 6

ABL09879



ABL09879 standard; cDNA; 3523 BP.  
 ABL09879;  
 26-MAR-2002 (first entry)  
 Drosophila melanogaster expressed polynucleotide SEQ ID NO 24119.  
 Drosophila; developmental biology; cell signalling; insecticide;  
 pharmaceutical; gene; ss.  
 Drosophila melanogaster.  
 WO200171042-A2.  
 27-SEP-2001.  
 23-MAR-2001; 2001WO-US09231.  
 23-MAR-2000; 2000US-191637P.  
 11-JUL-2000; 2000US-0614150.  
 (PEKE ) PE CORP NY.  
 Venter JC, Adams M, Li PWD, Myers EW;  
 P-PSDB; ABB65776.  
 New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
 Claim 1; SEQ ID NO 24119; 21pp + Sequence Listing; English.  
 The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).  
 The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 3523 BP; 910 A; 879 C; 911 G; 823 T; 0 other;

Query Match  
 Best Local Similarity 7.1%; Score 304.6; DB 23; Length 3523;  
 Matches 406; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 140 gggagagagtggaacgatattctcgaaggtgtttgttagcggtatccctccagacatt 199  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1717 ggcgatggcagctcgtctcccggaaggtatcgttggcgtcgcaccggacatc 1776  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 200 gatgaagatgagatcacagctagtcttcgtcttggccctctgattgtggatggcct 259  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1777 gatgagatgagattaccacttgcttcggcgttcggccattgctgcgattggcca 1836  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 260 cataaagctgagagaaacccattttctcctaaaggctatgcattcctcgtgttcaa 319  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1837 cacaagcggaatcccaagctgattttccgcccaggatgatgctcctcgtgttcag 1896  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 320 gatgaagctctgtcaggtctcattgatcatgcatgattgaagaagatggaaaactctac 379  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1897 gacgagcagtgacgagcagtaattgactgctgacacgagatgagacaagctgtat 1956  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 380 ctttgtatcaagtccactatcaaggataagccagtcagattcggccttgggaatctc 439  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1957 ctatgcgtttcttcgcgcagcatcaaggataaggcagatgacagattccttggcgctg 2016  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 440 agtgacagtgaacttgtgatggatgttccacagccacttgacccacgaaaaactatattt 499  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2017 gccgatgcggactagtgtgtgatgtctaccatgtcactggaccacgcacaaacggtgttt 2076  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 500 gttgggtgttctccgaccattacagctgtggagcttgagctggaatgataatgacgacta 559  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2077 gtggggcggtgccacgtctctgaagccttcgaactcggcaatgatcatgatgatgattg 2136  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 560 tacggaggtgtgtctacgctgggattgatccgaccctgagctgagctaaatacccaaaagga 619  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2137 tacggtggagtattgctatgtgaattgacaccgatccggaattaaagtatccaaagggc 2196  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 620 gctgggagagttgcgttctcttaacaacagagattacatgctgctatcagtcgcccgttt 679  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2197 gctggagctgtgtgacctctcgaatcagcagagctacatagcggccatctcagccagattt 2256  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 680 gtccagctgcagatggagagatagataaacgggt 714  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 2257 gtgcagctgcagatggcgatatagacaagcggt 2291  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7  
 ABL09813  
 ID ABL09813 standard; cDNA; 3597 BP.  
 XX  
 AC ABL09813;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23921.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 PI WPI; 2001-656860/75.  
 XX  
 DR P-PSDB; ABB65710.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
 XX  
 PS Claim 1; SEQ ID NO 23921; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).  
 XX  
 CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 3597 BP; 939 A; 907 C; 925 G; 826 T; 0 other;







```
DR WPI: 2000-097328/08.
XX DNA sequences preferentially expressed in IgA nephropathy patients,
PT proteins encoded by them, and antibodies to those proteins -
XX
XX Claim 1; Page 140; 180pp; Japanese.
XX This invention describes novel DNA sequences preferentially expressed in
CC IgA nephropathy patients, and DNA sequences stringently hybridizing to
CC them. Independent claims cover diagnostic reagents for IgA nephropathy
CC incorporating the antisense sequences; the treatment of IgA nephropathy
CC using the antisense sequences for mRNA inhibition; proteins associated
CC with IgA nephropathy, containing sequences encoded by the DNA sequences;
CC antibodies recognizing these proteins; the production of the proteins
CC by culture of host cells transformed with DNA encoding them; diagnostic
CC reagents for IgA nephropathy containing the antibodies; and compositions
CC for the treatment of IgA nephropathy which contain the antibodies. The
CC products of the invention can be used for the diagnosis and treatment of
CC IgA nephropathy. This sequence represents a PCR generated human IgA
CC nephropathy-associated protein INP303A cDNA fragment which is described
CC in the method of the invention.
XX
XX Sequence 305 BP; 95 A; 56 C; 54 G; 94 T; 6 other;
SQ
Query Match 7.0%; Score 299; DB 21; Length 305;
Best Local Similarity 98.0%; Pred. No. 3.6e-46;
Matches 299; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2797 acagagttccaggttgattataaccacacatgtgtacactataaatagaaccacga 2856
Db ACAGAGTTCAGGTTGTGATTATAACCCCAACATGTGTACACTATAAATAGAACCCACGA 246
QY 2857 gcagggttttcacagagtcagaatcttgtgacgcagtagtcaggcatcttcacaccg 2916
Db GCCAGGCTTTTACGACAGCTCAGAACTCTGTGACGCAGTAGTCAGGCATCTTCACACCG 186
QY 2917 acttgaatatgaagtgcagttgttggaacttggtgacatcttagttgtttgttttaa 2976
Db 185 ACTTGAAATATTGAAGTCAGTCTGTGTGGAACTTGGATCATCTAGTTGATTTGTTTAA 126
QY 2977 ttatgattccacatatgacaaaaatccagatccactaaataaaggaggtttatgtcta 3036
Db 125 TTATGATTCACATATACAAAAATCCAGATCCACTAAATTAATCAGGCTTTATGTCTCA 66
QY 3037 tgaataatccctgtgggttttaataataaattcttagtctaaacagttggttcactt 3096
Db 65 TGAATAANNNNNGTGGGTTTAAATCTCATAAACATTCTAGTCTAAACAGTTGGCTTCACCT 6
QY 3097 catga 3101
Db 5 CATGA 1
RESULT 13
AAD25036/c
ID RAA25036 standard; DNA; 11003 BP.
XX
AC RAA25036;
XX
XX 12-MAR-2002 (first entry)
XX Human oncostatin M (OSM) gene #1.
XX
XX Human; oncostatin M; OSM gene; haplotyping; genotyping; cancer;
KW lung inflammation; rheumatoid arthritis; chromosome 22q12.2;
XX single nucleotide polymorphism; SNP; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT variation replace (3536, A)
FT /*tag= a
```

```
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (3744, C)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (4062, A)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 4087..7003
FT /*tag= d
FT /product= "Human OSM protein"
FT 4087..4120
FT /*tag= e
FT /number= 1
FT 4120..5741
FT /*tag= f
FT replace (5691, G)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 5742..5884
FT /*tag= h
FT /number= 2
FT 5885..6421
FT /*tag= i
FT replace (5978, T)
FT /*tag= j
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6354, A)
FT /*tag= k
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6358, G)
FT /*tag= l
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT 6422..7003
FT /*tag= m
FT /number= 3
FT replace (6458, C)
FT /*tag= n
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6634, T)
FT /*tag= o
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6892, A)
FT /*tag= p
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (6995, T)
FT /*tag= q
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (7035, T)
FT /*tag= r
FT /standard_name= "Single nucleotide polymorphism (SNP)"
XX WO200187907-A2.
XX
XX 22-NOV-2001.
XX
XX 17-MAY-2001; 2001WO-US16157.
XX
XX 17-MAY-2000; 2000US-204868P.
XX (GENA-) GENAISSANCE PHARM INC.
XX Duda AE, Kazemi A, Koshy B;
XX
XX WPI: 2002-055680/07.
XX P-PSDB; AAE15318.
XX New isolated human oncostatin M polynucleotide, useful for therapeutic
XX purposes, for studying the expression and function of the
XX polynucleotide and for expressing oncostatin protein -
XX Example 1; Fig 1; 71pp; English.
XX The invention relates to genetic variants of human oncostatin M (OSM)
CC
```

CC gene. The invention also relates to compositions and methods for  
CC haplotyping and/or genotyping OSM gene in an individual. Polynucleotides  
CC of the invention are useful in studying the expression and function  
CC of OSM, and in expressing OSM protein for use in screening candidate  
CC drugs to treat diseases related to OSM activity. They are also useful  
CC for therapeutic purposes. Methods of the invention are useful for  
CC determining whether an individual has a haplotype or haplotype pairs.  
CC The method is also useful for improving the efficacy and reliability  
CC of several steps in the discovery and development of drugs for treating  
CC diseases associated with OSM activity, e.g. cancer, diseases involving  
CC lung inflammation and rheumatoid arthritis. The present sequence is  
CC human OSM gene located on chromosome 22q12.2.  
XX  
SQ Sequence 11003 BP; 2386 A; 3085 C; 3204 G; 2328 T; 0 other;

Query Match 5.3%; Score 225.2; DB 24; Length 11003;  
Best Local Similarity 60.8%; Pred. NO. 2.2e-32;  
Matches 405; Conservative 0; Mismatches 253; Indels 8; Gaps 2;

QY 1823 ttgtttgttgatgagtgagtcctcctcgtcaccagcgtgagtcagtggtacagtc 1882  
DB 1033 TTTTTTTTTTGAGACAGAGTCTAGCTCGTCACCCAGGCTGGAGTGGACGACATC 974  
QY 1883 tcagtcactcccaacctctgcctccgggttcaagtgtattcctcgttccttagcctccoga 1942  
DB 973 TCAGCTCACTACAACCTCTGCCACCTGGGTTCAAGCGCTTCCTCGCTTAGCCCTCCTGA 914  
QY 1943 gtagggtgaattacaggcaataacacacatgccacaaataatgttttatttttagtgag 2002  
DB 913 GTAGCTGGGATTTACAGCGCTGTGCCACCAAGCGCTAAATTTTGTATTTTAGTAGAG 854  
QY 2003 acaggttttccacgtgtgttgtagcaggccagtttcaactcctcagctcaaggatccact 2062  
DB 853 ACAGGTTTTCGCGATGTTGTCAGGCTGATCTCGAACTCTTAACCTCAGGTCATCCACT 794  
QY 2063 gctcaccctccctcaaaagtgtggattacaggtatgagccacccacctgcctggcctac 2122  
DB 793 GCCTCAGCTCCCAAGTGTGGGATTACAGGAGCGACCGCGCTTGGCTTATCTC 734  
QY 2123 ttggtttttttatgcacactaaaaatacctacatctcactgccttattccacaataagtt 2182  
DB 733 TTCTTATGAGGACACCATTCACATTGGATTAGGGCCCACTCACAATTTTACCTTA-ATTAT 675  
QY 2183 tcagagctgtgggtggttcattagaatactcagactgaattgttctcctgtgcaatgaa 2242  
DB 674 TCCTTTTTTTTCTTTTTCGAGACAAAGTCTTGCTGTGTTGCCAGGCTTAAGTGTACA 615  
QY 2243 atcctttgcccagtggttcattgtcactctgta-----gacattatggagcagcctagag 2295  
DB 614 GTGGTCTTTTCATAGCTCACTCAGCGCTTGAACCTCCGGGCTTAAGCGAGCGCTCCTGCT 555  
QY 2296 gccagaagcccagtgctctccttattgctgtctctcctggttctgtgacactcttcttc 2355  
DB 554 CAGCCTCCAGAGTAGCTGGGATTCACAGGTATGTGCTACACACCTGGCCAAATTTTACCTT 495  
QY 2356 tcctttttgacttt 2415  
DB 494 AATTACTTCTTTTTTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 435  
QY 2416 gtccaccaggtgagcacagaatcacatcatgactcactgcatgttcttctcttttg 2475  
DB 434 GTCAACCCAGCTGGAGTGCAGTGGCGCCAGCTTGGCTCAGTCAACCTCTGCTGCCAGG 375  
QY 2476 ttcatg 2481  
DB 374 TTCAAG 369

RESULT 14  
AAD25098/C  
ID AAD25098 standard; DNA; 11003 BP.  
XX

AC AAD25098;  
XX 12-MAR-2002 (first entry)  
XX Human oncostatin M (OSM) gene #2.  
DE Human; oncostatin M; OSM gene; haplotyping; genotyping; cancer;  
KW lung inflammation; rheumatoid arthritis; chromosome 22q12.2;  
KW polymorphism; ds.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FT misc\_feature 3536  
FT /\*tag= a  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 1"  
FT 3744  
FT /\*tag= b  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 2"  
FT 4062  
FT /\*tag= c  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 3"  
FT 5691  
FT /\*tag= d  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 4"  
FT 5978  
FT /\*tag= e  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 5"  
FT 6354  
FT /\*tag= f  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 6"  
FT 6358  
FT /\*tag= g  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 7"  
FT 6458  
FT /\*tag= h  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 8"  
FT 6634  
FT /\*tag= i  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 9"  
FT 6892  
FT /\*tag= j  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 10"  
FT 6995  
FT /\*tag= k  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 11"  
FT 7035  
FT /\*tag= l  
FT /note= "This degenerate base represents polymorphic  
FT site (PS) 12"  
XX WO200187907-A2.  
XX 22-NOV-2001.  
XX 17-MAY-2001; 2001WO-US16157.  
XX 17-MAY-2000; 2000US-204868P.  
XX (GENA-) GENAISSANCE PHARM INC.  
XX Duda AE, Kazemi A, Koshy B;  
PI

QY	2476	ttcatag 2481 
Db	374	TTCAG 369
RESULT 15		
ID	ABA74962	
AB	ABA74962 standard; DNA; 225 BP.	
XX	ABA74962;	
AC		
XX	01-FEB-2002 (first entry)	
XX		
DE	Human foetal liver single exon nucleic acid probe #23267.	
KW	Human; foetal liver; gene expression; single exon nucleic acid probe; ss.	
XX	Homo sapiens.	
XX		
PN	WO200157277-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	30-JAN-2001; 2001WO-USO0669.	
XX		
PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX	(MOLE-) MOLECULAR DYNAMICS INC.	
PA		
XX	Penn SG, Hanzel DK, Chen W, Rank DR;	
PI		
XX	WPI; 2001-483447/52.	
DR		
XX	Human genome-derived single exon nucleic acid probes useful for	
PT	analyzing gene expression in human fetal liver -	
PS	Claim 4; SEQ ID NO 23267; 639pp + sequence listing; English.	
XX	The invention relates to a single exon nucleic acid probe for	
CC	measuring human gene expression in a sample derived from human foetal	
CC	liver. The single exon nucleic acid probes may be used for predicting,	
CC	measuring and displaying gene expression in samples derived from human	
CC	fetal liver. The present sequence is a single exon nucleic acid	
CC	probe of the invention.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 225 BP; 44 A; 48 C; 66 G; 67 T; 0 other;	
Query Match	5.3%; Score 225; DB 22; Length 225;	
Best Local Similarity	100.0%; Pred. No. 1.5e-32;	
Matches 225; Conservative	0; Mismatches 0; Indels 0; Gaps	
QY	3353 gfggaagttaagccatattgtcttgatgatcagctgtgtgatgaatgcagggggcccg 3412	
Db	1 gfggaagttaagccatattgtcttgatgatcagctgtgtgatgaatgcagggggcccg 60	
QY	3413 tgtggggggaaatttgcctcaattttctgtgctaattgttacctgtctgcagttacttgt 3472	
Db	61 tgtggggggaaatttgcctcaattttctgtgctaattgttacctgtctgcagttacttgt 120	
QY	3473 gaatatcttggtctgctacattctgtgctggcagggaattccacaagcccctgg 3532	
Db	121 gaatatcttggtctgctacattctgtgctggcagggaattccacaagcccctgg 180	



Qy 3533 aaggaaggcgggtgaccgccctcggcatatttcattccgctggaac 3577  
|||||  
Db 181 aaggaaggcgggtgaccgccctcggcatatttcattccgctggaac 225  
|||||

Search completed: September 1, 2002, 09:35:35  
Job time: 9746 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2002, 06:51:40 ; Search time 95.55 seconds  
(without alignments)  
10992.453 Million cell updates/sec

Title: US-09-730-559b-1

Perfect score: 4276

Sequence: 1 ttctaccgttttttcctgc.....atcagaaaaa4276

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2.6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2.6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2.6/prodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2.6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	216.8	5.1	17327	1	US-07-906-871-15
2	213.8	5.0	2061	2	US-08-960-022-11
3	212.6	5.0	14747	4	US-09-608-285A-42
4	212.6	5.0	15977	4	US-09-608-285A-59
C 5	210.8	4.9	2713	2	US-08-916-901-6
C 6	210.8	4.9	2713	4	US-09-154-602-6
7	209.8	4.9	2839	4	US-09-061-702-1
8	208.2	4.9	53526	3	US-08-658-136-2
9	208.2	4.9	53577	3	US-08-658-136-1
C 10	207.6	4.9	246240	2	US-08-724-394A-20
C 11	207.6	4.9	246240	2	US-08-724-394A-21
C 12	207.6	4.9	246240	2	US-08-724-394A-22
13	206.8	4.8	1442	2	US-08-454-557C-120
14	206.8	4.8	1442	2	US-08-340-436D-120
15	206.8	4.8	1442	2	US-08-450-673C-120
16	206.6	4.8	246240	2	US-08-724-394A-20
17	206.6	4.8	246240	2	US-08-724-394A-21
18	206.6	4.8	246240	2	US-08-724-394A-22
19	206.4	4.8	1701	4	US-09-078-294-9
20	206.2	4.8	1988	2	US-08-257-963B-11
21	206.2	4.8	1988	4	US-08-367-841A-11
22	206.2	4.8	1988	5	PCT-US95-07201-11
23	206.2	4.8	22481	4	US-08-367-841A-43
24	206.2	4.8	22481	5	PCT-US95-07201-43
25	205.4	4.8	59065	4	US-09-813-817-3
26	205.2	4.8	646	4	US-09-385-982-314
27	203.6	4.8	72928	3	US-09-009-913-1

C 28 202.6 4.7 14636 4 US-09-173-914-6 Sequence 6, Appli  
29 202.2 4.7 31571 1 US-08-323-443B-1 Sequence 1, Appli  
C 30 202 4.7 36741 4 US-09-301-665-3 Sequence 3, Appli  
31 201.8 4.7 619 4 US-09-385-982-358 Sequence 358, App  
C 32 201.6 4.7 2115 1 US-08-395-800A-7 Sequence 7, Appli  
33 200.6 4.7 4823 2 US-08-457-254-5 Sequence 5, Appli  
34 200.6 4.7 4823 2 US-08-484-257-20 Sequence 20, Appli  
35 200.6 4.7 4823 3 US-08-999-927-5 Sequence 5, Appli  
36 200.6 4.7 4823 4 US-08-461-819-5 Sequence 5, Appli  
37 200.6 4.7 4823 5 PCT-US94-08806-28 Sequence 28, Appli  
38 200.6 4.7 4823 5 PCT-US95-01829-5 Sequence 5, Appli  
39 200.6 4.7 4823 5 PCT-US95-16626-5 Sequence 5, Appli  
C 40 200 4.7 2127 1 US-08-832-883-54 Sequence 54, Appli  
C 41 200 4.7 2127 2 US-08-832-877-54 Sequence 54, Appli  
C 42 199.6 4.7 1386 2 US-08-687-080-76 Sequence 76, Appli  
43 199.4 4.7 7676 1 US-08-451-777A-7 Sequence 7, Appli  
44 199.4 4.7 7676 2 US-08-451-778A-7 Sequence 7, Appli  
45 199.4 4.7 7676 2 US-08-998-208-7 Sequence 7, Appli

## ALIGNMENTS

RESULT 1

US-07-906-871-15/c

; Sequence 15, Application US/07906871

; Patent No. 5340739

; GENERAL INFORMATION:

; APPLICANT: Stevens, Richard L.

; APPLICANT: Avraham, Shalom

; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1225 Connecticut Avenue, N.W., Suite 300

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/906,871

; FILING DATE: 19920103

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/816,289

; FILING DATE: 03 JAN 1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/635,544

; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US89/03051

; FILING DATE: 13-JUL-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/224,035

; FILING DATE: 13-JUL-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Cimbal, Michele A.

; REGISTRATION NUMBER: 33,851

; REFERENCE/DOCKET NUMBER: 0627.2830004

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)833-7533

; TELEFAX: (202)833-8716

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 17327 base pairs

; TYPE: NUCLEIC ACID







```

Db 1241 TATTTTATGAGACAGAGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACC 1182
QY 2048 tcaaggatccacctccacctcaaaagtctgggattacaggtacaggtacaggtacaggtac 2107
Db 1181 TAGGTGATCCACCCGCTCGGCTCCCAAAAGTGTGGGATTACAGGATGAGCCACGCG 1122
QY 2108 ceatgctgctacttggtgtttttatgcacactaaaaataacc 2151
Db 1121 GCCTGGCGGACACCATTTCTTTTCAATGAAACAGTCTAATTCC 1078

RESULT 7
US-09-061-702-1
; Sequence 1, Application US/09061702
; Patent No. 6165737
; GENERAL INFORMATION:
; APPLICANT: Wang, Xiaodong
; APPLICANT: Liu, Xuesong
; TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN
; TITLE OF INVENTION: APOPTOSIS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/061,702
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMillian, Nabeela R.
; REGISTRATION NUMBER: P-43,363
; REFERENCE/DOCKET NUMBER: UTSD:546
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512)418-3000
; TELEFAX: (512)474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-061-702-1

Query Match 4.9%; Score 209.8; DB 4; Length 2839;
Best Local Similarity 76.0%; Pred. No. 1.6e-34;
Matches 259; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1823 ttgtttgttgagtggaagtctctccgtccaccagctgagtgagtgagtgagtc 1882
Db 1984 TTTGTTGTTGAGACAGGGTCTCACTCTGCTCGCTGAGGTGAGTGAATGCAATC 2043
QY 1883 tcagctcaactcccaacctctgcctcccggttcaagtgtattctctccttagcctccga 1942
Db 2044 TTGGCTCACTGCAACCTCCGCCCTCTCAGGTTCAAGTGGTTCCTCGCTCAGCTCCCAA 2103
QY 1943 gtagggtgaattacaggtacatataccaccatgccccaaatgtttgtatttttagtggag 2002
Db 2104 GTAGCTGGGATTACAGGCACCTACCACCGGGCAGCTAATTTTGTATGTTTAGTAGTA 2163
QY 2003 acaggggttcacgtattgttcagggcagtttcaaaactcctacactcaaggaatccact 2062
Db 2164 ACGGGGTTTCCACCATGTTGGCCAGGGTGTCTTCGAACTCCTGACCTCAGGTGATCGCCC 2223
```

```

QY 2063 gctcaccacctcaaaagtctgggattacaggtacaggtacaggtacaggtacaggtac 2122
Db 2224 GCCTCGGTCTCCCAAAAGTGTGGGATTACAGGCATGAGCCACTGCACCTGACCTGCTGAA 2283
QY 2123 ttggttttttatgcacactaaaaataaccatctacatctcactg 2163
Db 2284 TTGTTTATAATGCGAAGAAATAGGAAACCCCAATGCTCTG 2324

RESULT 8
US-08-658-136-2
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53526 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match 4.9%; Score 208.2; DB 3; Length 53526;
Best Local Similarity 75.7%; Pred. No. 7.2e-34;
Matches 258; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1818 cctactttgtttgttgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 1877
Db 989 CTTTCTTTTCTTTTGTGAGATGGAGTCTCACTCTGTACCCAGGCTGGAGTGCAGTGGG 1048
QY 1878 cagtcagctcactcccaacctctgcctcccggttcaagtgtattctcctccttagcct 1937
Db 1049 TAACTCAGGTGACACGCGACCTCCGCCCTCCCGGGTTCACCGTGTCTCCTGCCCTCAGCCT 1108
QY 1938 cccgagtgagtggaattacaggtacatataccaccatgcccaacaaatgtttgtattttag 1997
Db 1109 CCCGAGTAGTGGGATTACAGGCCACCCACCATGGCTGGCTAATTTTGTATTATTAG 1168
```

Qy 1998 tggagacagggtttccacctggttgcagcgagctttcaaacactccttgacctcaaggatc 2057  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1169 TAGAGACGGGGTTTTTCCACAGTTGGCCAGTTGGTCTCGAACTCTTGCCCTCATGTGACC 1228  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 2058 cactgcctcacccccctcaaagtgcgtggattaccagggcatgagccacccacatgctcg 2117  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1229 CGCCTGCCTGGCCTCCCAAGTCTGGGATTACAGGTGTGAGCCACTGTGCCITGGCCTG 1288  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 2118 cctaactgttttttatgcacacataaaaaatacctcacatct 2158  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1289 GCTTTCCTGTGTTCTTTTCTCTCCTCTTAGTTTCCOCCCTTT 1329  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9  
US-08-658-136-1  
; Sequence 1, Application US/08658136  
; Patent No. 6071717  
; GENERAL INFORMATION:  
; APPLICANT: KLINGER, KATHERINE W  
; APPLICANT: LANDES, GREGORY M  
; APPLICANT: BURN, TIMOTHY C  
; APPLICANT: CONNORS, TIMOTHY D  
; APPLICANT: DACKOWSKI, WILLIAM  
; APPLICANT: GERMINO, GREGORY  
; APPLICANT: QIAN, FENG  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: ONE MOUNTAIN ROAD  
; CITY: FRAMINGHAM  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/658,136  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LASSEN, ELIZABETH  
; REGISTRATION NUMBER: 31,845  
; REFERENCE/DOCKET NUMBER: GEN4-17.8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 508-872-8400  
; TELEFAX: 508-872-5415  
; INFORMATION FOR SEQ ID NO. 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-658-136-1









---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2002, 06:17:41 ; Search time 3375.57 Seconds  
(without alignments)  
17097.255 Million cell updates/sec

Title: US-09-730-559B-1  
Perfect score: 4276  
Sequence: 1 ttctaccgttttttcctgc.....atcagaaaaaa 4276

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	605.4	14.2	1077	11 AK015381	Mus muscu
2	605.4	14.2	1586	11 AK015401	Mus muscu
3	598.2	14.0	860	10 BM456398	AGENCOURT
C 4	507	11.9	529	9 AA772278	ai42d03.s
C 5	495	11.6	506	9 AI392674	tg47c02.x
C 6	490	11.5	501	10 B8468016	hz77c06.x
7	481	11.2	492	10 B504477	hz57a01.x
8	447.6	10.5	1066	11 AK021394	Mus muscu
C 9	440.2	10.3	465	10 B502228	hy14e07.x
C 10	423.2	9.9	559	12 AQ883641	HS_5478_B
C 11	420.2	9.8	486	10 BF903366	QV1-MT022
C 12	410.6	9.6	425	9 AW237641	xm74h08.x
C 13	408.4	9.6	433	9 AI962924	wt24h06.x
C 14	399.6	9.3	438	10 H71226	ys12e09.s1
C 15	383.8	9.0	395	9 AI802170	tx25b02.x
16	376	8.8	398	10 H71225	ys12e09.r1
17	363.8	8.5	650	9 BB612073	BB612073

18	357.2	8.4	636	10 BE302536	ba67b01.y
19	354.2	8.3	399	10 BM286426	526494 MA
C 20	353.8	8.3	365	9 AI093858	qa30q05.s
C 21	353	8.3	364	9 AA634469	zu76a06.s
22	345.4	8.1	661	10 BG081288	H3063D07-
C 23	343.2	8.0	465	10 BF092226	PM2-TN009
C 24	337.8	7.9	548	9 AW958536	EST370606
C 25	326.4	7.6	329	10 Z19960	HSAABDCI B
C 26	323.8	7.6	914	10 BI648114	603278431
C 27	320.8	7.5	822	10 BG068252	H3063D07-
C 28	307	7.2	385	10 T93967	YD56d10.s1
29	302	7.1	302	9 AA381126	EST94221
30	299.2	7.0	820	10 BF181733	601805546
31	291.4	6.8	505	10 BE967640	601648932
32	290.2	6.8	410	10 H57530	YR05b10.r1
33	281.4	6.6	634	10 BE383195	601298746
34	281	6.6	578	10 BI359798	384502 MA
35	278.8	6.5	623	9 AV282552	AV282552
C 36	274.2	6.4	521	9 AW653337	101985 MA
37	263.4	6.2	777	12 BHI26248	GG007 CH1
38	258.8	6.1	372	9 AW430013	68737 MAR
C 39	257.2	6.0	438	9 AU022414	AU022414
40	256.8	6.0	661	9 BB405049	BB405049
C 41	252	5.9	321	10 BG956024	PM2-CT080
42	248.6	5.8	593	10 BE977433	bs63h12.y
C 43	245.6	5.7	274	9 AA782211	ai47h12.s
44	239.6	5.6	582	10 BG306825	fm07e09.y
45	236.4	5.5	499	10 BI349383	dac60e03.

#### ALIGNMENTS

#### RESULT 1

AK015381 1077 bp mRNA linear HTC 19-JAN-2002  
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930444G21:RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) containing protein, full insert sequence.  
DEFINITION  
ACCESSION AK015381 GI:12853699  
VERSION HTC; CAP trapper.  
KEYWORDS Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library  
SOURCE clone:4930444G21.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Carninci,P. and Hayashizaki,Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 (sites)  
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3 (sites)  
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)



## TITLE

High-efficiency full-length cDNA cloning

## JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

## MEDLINE

99279253

## PUBMED

10349636

## REFERENCE

2 (sites)

## AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

## TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

## JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

## MEDLINE

20499374

## PUBMED

11042159

## REFERENCE

3 (sites)

## AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, K., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

## TITLE

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

## JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

## MEDLINE

20530913

## PUBMED

11076861

## REFERENCE

4 (sites)

## AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

## TITLE

Functional annotation of a full-length mouse cDNA collection

## JOURNAL

Nature 409, 685-690 (2001)

## MEDLINE

5 (bases 1 to 1586)

## PUBMED

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanegaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saigo, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

## TITLE

Direct Submission

## JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCTCGAGTCAAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTCAAGTAAATTAATCCCTCCCTCCCTCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

## FEATURES

## Source

Location/Qualifiers

1..1586

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="MGD:MGI:1897157"

/db\_xref="taxon:10090"

/clone="4930447D24"

/sex="male"

/tissue\_type="testis"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

34..921

/note="RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) containing protein data source: Pfam, source key: PF00076, evidence: ISS

putative"

/codon\_start=1

/protein\_id="BAB29832.1"

/db\_xref="GI:12853731"

/translation="MEDGFLDDGGDQPLHSGSPHCFTHQNGRVERYSRKVFVGG LPDIDDEITASFRFGLIVDPHKAESKSYFPKPKYAFLLFQDESSVQALIDACI EEDGKLYLCVSPTIKDKPVQIRPWLSDSDVDGSDQPLDPRKTFIVGVGPRPLRAV ELAMIMRLYGVGYAGIDPDPELKPKGARRVAFSNQOSYIAAISARFVQLQHGIED KRYVEKPYVLDLQDCQACGCGGKFAFFCANVTCLOYYCEYCNAAIHSRAGREHF KPLVKEGGDRPHISFRWN"

1568..1573

/note="putative"

1586

/note="putative"

BASE COUNT 426 a 329 c 370 g 461 t

## ORIGIN

Query Match 14.2%; Score 605.4; DB 11; Length 1586;  
Best Local Similarity 91.3%; Pred. No. 3.1e-66;  
Matches 642; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
Qy 30 caggtcagcttccactgtttccgagtggaagatgatttcttgatgagtcggcgtgggagtc 89  
Db 11 CTGGTCAGCTTCATTTGTTCCGATGGAAGATGGATCTCTGGATGATGGCGTGGGATC 70  
Qy 90 agcctcttcacatagtgccgtgggtggtccactcactcactcactcactcactcactcactc 149  
Db 71 AACCTCTTCATAGTGGTCTGGGGTCACCTCACTGCTTCACTCACCAGAAATGGGAGAG 130  
Qy 150 tggaaacgatatctcgaagagtggtttgtagcggtatgcctccacagacattgatgaagatg 209  
Db 131 TGGACGAGATCTCTCGCAAGAGTGTTGTGGGGTGGATGGCTTCTTGATATGATGAAG 190  
Qy 210 agatcacagctagttttctgcgctttggccctctgattggtgattggcctcctcctcctg 269  
Db 191 AGATCACAGCTAGTTTCCGTCGCTTTGGCCCTTTGATTTGATTTGGCTTCATAAGCAG 250  
Qy 270 agagcaaatcctatttctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 329  
Db 251 AGAGCAAACTTATTTTCCACCAAAAGGCTATGATCTCTGCTCTCTTCAAGATGAAAGTT 310  
Qy 330 ctgtgcagctctcattgatcattgatcattgaagatggaaactctacacctcttctgtat 389  
Db 311 CTGTTTCAGCTCTCATTTGATGCATGATTTGAAGAGATGGAAACTTTACCTGTGTAT 370  
Qy 390 caagtcacctcactcaagagataagccagtcacagattcggccttggaactcagtcagtcagtc 449  
Db 371 CAAGTCCAAACCATCAAGGATAAACAGTGCAGATCCGCGCCCTCGGAATCTCAGTCACAGTG 430  
Qy 450 actttgtgatgagtggttcaacagcacttgaccacacacacacacacacacacacacacac 509  
Db 431 ACTTTGTGATGGTGGCTCACAGCCACTTGGACCCAGGAAACAAATATTTTGTGGTGGTG 490  
Qy 510 ttcttcgaccttaccagctgtggagcttgcgattggaatgattcggtcgtacacagagtg 569  
Db 491 TTCTTCGACCAATACGAGCTGTGGAGCTTGCAATGATTAATGATCGGCTGTATCGAGCG 550  
Qy 570 tgtgtacgctgggattgataccgacctgagctaaataaccacacacacacacacacacacac 629  
Db 551 TCTGCTATGCTGGAATCGATCTACGCCAGAGCTCAAAATACCCAAAGGAGCTAGAAAGAG 610





) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCGAAGTGGGAGCGCGCCACCAATTTTTTTTTTTTTTTT  
TTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH.

BASE COUNT 196 a 80 c 86 g 167 t  
ORIGIN

Query Match 11.9%; Score 507; DB 9; Length 529;  
Best Local Similarity 98.7%; Pred. No. 8.1e-54;  
Matches 522; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 3740 gtgagattctgtgtaactcaacaatatgaactccttttttgcattgcccacgtggtgc 3799  
|||||  
Db 529 GTGAGATCTGTGTCACTGCAACAATAATGAACCTCTTTTCGTATTGCCATCGGGTGC 470  
QY 3800 atggaagtattctcttctgttctggaaccagagatcccaacttctctgcaacat 3859  
|||||  
Db 469 ATGGAAGTTTATCTCTGTGTTTCTGCGAACCAGAGATCCAAACTTCTCGCAACAT 410  
QY 3860 ttcttagaggagagagaaatattaaagagaaatgaacaatagagtattttgggtt 3919  
|||||  
Db 409 TTTCTTAGAGGAGAGAGAAATATTAAGAGAAATGAACAATAGAGTATTITGGGTT 350  
QY 3920 tttaattaaattattgtaataataacataagaatactttttattataaataaccatg 3979  
|||||  
Db 349 TTTAATTAATTTATTGTTAATAATAACATATAAGATCTTTTATTAAATAACCAATG 290  
QY 3980 caacataacacatctcgtctctatctgacagtttttccccagggagtgctttgctttt 4039  
|||||  
Db 289 CAACAATAACACTATCGGTCTATCTGACAGTTTTCGCCAGGGAGTGCATTTCGCCTT 230  
QY 4040 tctcttcttt 4097  
Db 229 TCTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 170  
QY 4098 ttaatttttttaacgaatgagagaaatgaacaatttttaataagaacagatgttaga 4157  
Db 169 TTAATTTTAAACGAATGAGGAAGTTAACAAATTTTAAATGAAAGAGCATGTTAGA 110  
QY 4158 gcaacaaatgcataagcaagactgagcagcattataataattttcagggttttgaggc 4217  
Db 109 GCAACAATAATGCATAAGCAAGACTGAGCAGCATTTATAATTTTTCAGGCTTTGAGGC 50  
QY 4218 tgaacataatttcattatccctcctcaaaaagtaccaccacacagaataa 4266  
Db 49 TGAACATAATTTTCATATCCCTCAAAAAGTTACCACCACATCAGAAAAA 1

RESULT 5  
LOCUS AI392674/c 506 bp mRNA linear EST 04-FEB-1999  
DEFINITION t947c02.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2111906 3', mRNA sequence.  
ACCESSION AI392674  
VERSION AI392674.1 GI:422221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 506)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 444.  
Location/Qualifiers  
1. 506  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2111906"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbH19W, testis NHT, and B-cell  
NCI-CGAP-GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
728408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
185 a 75 c 82 g 164 t

FEATURES  
source

BASE COUNT  
ORIGIN

Query Match 11.6%; Score 495; DB 9; Length 506;  
Best Local Similarity 99.8%; Pred. No. 2.5e-52;  
Matches 506; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3762 acaatatgaactccttttcgtattgccatcggtgtgcagtggaagttttattctctgtt 3821  
Db 506 ACAATATGAACCTCTTTTTCGTATTCGCATCGGGTTCATCGGAAGTTTATTCTCTGTT 447  
QY 3822 ttgctgaaaccaagagatcccaacttcctgcaacatttcttagaggagagagaaa 3881  
Db 446 TTGCTGGAACCAAGAGAGATCCAAACTTCTCGCAACATTTCTTAGAGGAGAGAGAAA 387  
QY 3882 tattaaagagaatgaacaatagagtatttgggtttttaataataattattgttaata 3941  
Db 386 TATTAAAGAGAAATGAACAATAGAGTATTTTGGGTTTAAATTAATTAATTTGTTAATA 327  
QY 3942 atataacataaagaaacttttattataaataaccatgcacaataaacacatcgtctca 4001  
Db 326 ATATAACATATAAGATACCTTTTATTAAATACCATGCAACAATAACACTATCGGTCTA 267  
QY 4002 tctgacagtttttccccagggagtgcttttgccttttctcttcttttttttttttttt 4061  
Db 266 TCTGACAGTTTTCCTCCAGGAGTCTTTTGCCTTTTCTTTCTTTTCTTTTCTTTT 207  
QY 4062 catctttttgttctctctcttttttccatcccttttttaatttttttaacagcaatggag 4121  
Db 206 CATCTTTTGTG-TCTCTCTCTTTTTCCTATCCCTTTTAAATTTTAAACGAATGGAG 148  
QY 4122 gaagtttaacaatttttaaggaagagcatgttagagcaacaaaatgcataagcaagact 4181  
Db 147 GAAGTTAACAATTTTAAATGGAAGAGCATGTTAGAGCAACAAATGTCATAGCAAGACT 88  
QY 4182 gacgacattataataattttcagggttttgagggtgaacataatttctattatccctca 4241  
Db 87 GAGCAGCATTTATAATTAATTTTCAGGGTTTGTAGGCTGAACATAAATTTTATTATTCCTCA 28

RESULT 6

BE468016	BE468016	501 bp	mRNA	linear	EST 27-JUL-2000
LOCUS	hz77c06.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3213994 3'				
DEFINITION	mRNA sequence.				
ACCESSION	BE468016				
VERSION	BE468016.1 GI:9513791				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 501)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
JOURNAL	Tumor Gene Index				
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbsr@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LUNL, send email to: info@image.llnl.gov Seq primer: -40UP from Gibco High quality sequence stop: 464.				
FEATURES	Location/Qualifiers				
source	1..501				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone_image="IMAGE:3213994"				
	/clone_lib="NCI_CGAP Lu24"				
	/tissue_type="carcinoid"				
	/lab_host="DH108"				
	/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP-Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439).. Subtraction by Bento Soares and M. Fatima Bonaldo."				
BASE COUNT	162 a 84 c 90 g 165 t				
ORIGIN					
	Query Match 11.5%; Score 490; DB 10; Length 501;				
	Best Local Similarity 99.8%; Pred. No. 1.le-51;				
	Matches 501; Conservative 0; Mismatches 0; Indels 1; Gaps 1;				
Qy	2797 acaggtccaggtttgattataaccaccaatggtcacactaaatagaaccacga 2856				
Db	1 ACAGGTCCAGGTTTGATTATTAACCCACACTGTGTACACTATAATAGAACCACGA 60				
Qy	2857 gccaggcttttacgcagagctcagaatcttgtgcgcagtagtcaggcatcttcacaccg 2916				
Db	61 GCCAGGCTTTTTCACACAGCTCAGAACTTCTGCACGCACTAGTCAGGCATCTTCACACGG 120				
Qy	2917 actgaatatgaagtcaggttgtggaaacctbggatcattagtgtattttttaaaa 2976				
Db	121 ACTTGAATATTGAAGTCAGTTGTGGAACTTGGATCATCTTAGTTGATTTGTTTAAA 180				
Qy	2977 ttatgattcccacatgatgacaaaaatccagatccactaatataaacaggggttatgtcta 3036				
Db	181 TTATGATTCACATATGACAAAAATCCAGATCCACTAATTAATATGAGGGTTTATGTCTA 240				
Qy	3037 tgaataatctcctgbggttttaataatcacaattctagtctaaacagttcggttcactt 3096				
Db	241 TGAATAATCTCCTGTGGGTTTTAATCTATAACATTCTAGCTAAACAGTTGGCTTCACTT 300				

sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4 (sites)  
THE RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5 (bases 1 to 1066)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,  
Arakaki, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,  
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,  
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,  
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,  
Kurthara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,  
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,  
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I.,  
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and  
Hayashizaki, Y.  
Direct Submission  
Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 10.0 and subtraction to  
Rot = 239.4. Second strand cDNA was prepared with the primer  
adaptor of sequence [5'  
GAGAGAGATCTCGATTGAATTAATTCCTCCCTCCCTCC 3']. cDNA was cleaved  
with BamHI and XhoI. Vector: a modified pBluescript KS(+) after  
bulk excision from Lambda FLC I. Cloning sites, 5' end: Sall; 3'  
end: BamHI. Host: DH10B.  
FEATURES  
Source  
1..1066  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="MGD:MG1:1899445"  
/db\_xref="taxon:10090"  
/clone="E13011861e"  
/tissue\_type="eyeball"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="0 day neonate"  
<1..427  
/notes="RNA recognition motif. (a.k.a. RRM, RBD, or RNP  
domain) containing protein  
data source:Pfam, source key:PF00076, evidence:ISS  
putative"  
/codon\_start=2  
/protein\_id="BAB32395.1"  
/db\_xref="GI:1286272"  
/translation="RLRAVELAMIMDRLYGVGVAGIDTDELKPKGAGVAFPSNQ  
QSVIAISARFVQLQGEHDIKRVKPYVLDQLCECGAGCGKFAFFCNAVYLC  
QYTCYCAAIHSRAGREHKPLVDGGDRPHISFRWN"  
291 a 221 c 238 g 316 t  
BASE COUNT  
ORIGIN

Query Match		10.5%;	Score 447.6;	DB 11;	Length 1066;
Best Local Similarity		79.7%;	Pred. No. 1.2e-46;		
Matches 719;		Conservative 0;	Mismatches 129;	Indels 54;	Gaps 14;
QY	3350	cagggtgaagttaagccaatgtcttgatgcacgtgtgtgatgaatgtcaggggcc	3409		
Db	197	CGGTGGAGGTTAAGCCATATGCTTGGATGACGAGCTGTGTGATGAATGTCAAGGGGCC	256		
QY	3410	cgttggtgggggaaattgtccattttctgtgctaaatgtttacctgtctgcagttattac	3469		
Db	257	CGTGTGGGGGAAATTTGCTCCATTTTCTGTGCTAAATGTACTGTCTGCAGTATTAC	316		
QY	3470	tgtgaattattgctgggctgtaccattctcgtgctggcagggaattccacaagccctg	3529		
Db	317	TGTGNATATTGCTGGGCTGCTATTCACTCTCGTGTGGCAGAGAAATCCACAGCCCTTG	376		
QY	3530	gtgaaggaaggcgtgtacccctcgtgcataatttcacgtcgtgaaact-----aaagat	3585		
Db	377	GTGAAGGAAGGTGGTACCCTCCGCTATTTTTCATTCGCTGGAACCTGAAGAATGAAT	436		
QY	3586	aactgcagtgctcatittcaggcctcagaataaagtgaactcttctgttcaattctgcaccc	3645		
Db	437	GACTCCAGAGCTCATTTGTCAGGCTCAGAGAAGTGCCTCTTTTGTAAATTC-CAACTC	495		
QY	3646	ttctcaacctcttcacgtggtcatgtcctttttagcagctgtgaacttaactatagta	3705		
Db	496	TTCTCTAGCCTCTTCACACTGGCATGTCC-TGCACAGCAGTGTGTAAC-TATCAGTGGTA	553		
QY	3706	taatgaagaagaatgacctataataagggtttttgttagattcttctgtcactgcacaaaca	3765		
Db	554	GAATGAAGAAGAACCACTA-CACATAGGTATTTTGTAGACTCTCGTGTACATGCAACAG	612		
QY	3766	tatgaactctct--ttttcgtattgcatcgggttgcatggaagtctt-attctctgttt	3822		
Db	613	TATGTTCAACTCCATTTCACATTTGCCATATATGTTGATCCAGGTTTAAATTCCTCTGTT	672		
QY	3823	tgtggaacccaagagagatccaaactcctgcacatttctcttagaggagagagaaat	3882		
Db	673	TGCTGAGGCCAAGAGGATCTAACTTGCTGTAGCATTTTCTTAGAG----GAGAGAAAT	728		
QY	3883	attaaagagaatgaacaatagagtatttgggtttttaaattatgtttaaataa	3942		
Db	729	ATTAACAGAGAAATGAGACATA-----TTTTGAGGTTTAAATGCTGGTAATGA	778		
QY	3943	tataacataagaatacattttattataaataaacatgcaacaataaacactatcggtctat	4002		
Db	779	TATAACATATAAGAAACTTTATTGAGATAATC--ACCAACAGTAACACTATCATCTGTT	836		
QY	4003	ctgacagttttccccccagggaagtccttttgcctttctcttctcttcttcttcttctt	4062		
Db	837	CTGACACAGTCTCCCCCAGGAAGTGTCTTGCCAGTCTTTTCTT-----882			
QY	4063	atctttttgttctctctctttttccatcccttttttaatttttttaacagcaatggagg	4122		
Db	883	---TCATTCTCTCTCTCTCTTTTCTATCCCTTTTA--TTCTTTAACAGCAATGGAGG	937		
QY	4123	aagttaacaatttttaa-tggaaagagcatgttagagcaacaacaaatgcataagcagaact	4181		
Db	938	AAGTTAACAGTATTATAACAGGAAGAGTATGTCAGAGCAAGCAAAATGCATGACGAAT	997		
QY	4182	gagcagcattataataattttcagggttttgaggctgaacataaatttcattatccctca	4241		
Db	998	GAGCAG-----TGTAACCATTAAGAGCTTTGAGGCTGATGATGACTGTGATGCTGCCA	1050		
QY	4242	aa 4243			
Db	1051	AA 1052			

RESULT 9  
BE502228/c 465 bp mRNA llinear EST 04-AUG-2000  
LOCUS BE502228

DEFINITION	hvl14e07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3197316 3', mRNA sequence.
ACCESSION	BE502228
VERSION	BE502228.1
KEYWORDS	GI:9704636
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1. (Bases 1 to 465)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov Seq primer: -40UP from Gibco High quality sequence stop: 455. Location/Qualifiers 1. .465 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3197316" /clone_lib="NCI_CGAP_GC6" /tissue_type="pooled germ cell tumors" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469084-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	174 a 69 c 74 g 148 t
ORIGIN	
Query Match	10.3%;
Best Local Similarity	98.1%;
Matches 456;	Conservative 0;
Mismatches 8;	Indels 1;
Gaps 1;	

QY	3798	gcattggaagtttattctcttctgttggaacacagagatccaaacttcttccaac	3857
Db	465	GCATGGAAGTTTATTTCTGTGTTTCTGGAACCAAGAGGATCCAACTCTCTGCCAC	406
QY	3858	atttcttagagagagagaaattataaagagaatgaacaatagatgattttggg	3917
Db	405	ATTTTCTTAGAGGAGAGAGAAATATTAAGAGAAATCAACAATAGAGTATTTGGG	346
QY	3918	tttttaataaattattgtttaataataacataataaactttttataaataacca	3977
Db	345	TTTTTAATTAATTTGTTAATAATAACATATAAGAAATACCTTTTATAAATAACCA	286
QY	3978	tgaacaataacactatcggtctatctgcagtttttccccccaggaagcttttgccct	4037
Db	285	TGCAACAATAACACTATCGCTCTATCTGACAGTTTTCCTCCAGGGAAGTCTTTGCCCT	226
QY	4038	tttcttcttttttttttttttttttctt- tttttgttctctcttttttccatccctt	4096
Db	225	TTTCCTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCCATCCCTT	166
QY	4097	tttaatttttttaacagcaatggaggaagttaacaattttttaatggaaagagcatgttag	4156

```

165 TTTAAATTTTACAGCAATGGAGAGTTTAAACAATTTTATGGAAGAGCATGTAG 106
|||||
4157 agcaacaaatgcataagcaagactgagcagcattataataattttcagggttttgag 4216
|||||
105 AGCAACAAATGCATAGCAAGACTGAGCAGCATATATAATTAATTTTCAGGGTTTTCAGG 46
|||||
4217 ctgaacataatttcattatccctcaaaaagtaccacacacacag 4261
|||||
45 CTGAACATAATTTTATTATTCCTCAAAAGATTACCACACATCAG 1

RESULT 10
AQ883641/c 559 bp DNA linear GSS 09-NOV-1999
LOCUS HS_5478_B2_C11_77A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=9246 Col=22 Row=F, DNA sequence.
ACCESSION AQ883641
VERSION AQ883641.1 GI:6315108
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 559)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 9246 Row: F Column: 22
Seq primer: T7
Class: BAC ends
High quality sequence stop: 559.
Location/Qualifiers
l..559
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=9246 Col=22 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 179 a 124 c 90 g 161 t 5 others
ORIGIN

Query Match 9.9%; Score 423.2; DB 12; Length 559;
Best Local Similarity 92.9%; Pred. No. 1.8e-43;
Matches 474; Conservative 0; Mismatches 32; Indels 4; Gaps 3;

Qy 2997 aaatccagatccactaattaaatgaggtttatgtctatgaataatctcctgtgggtt 3056
|||||
Db 551 AAAACTCAGACTCACTAAT--AAATAGAGGTTTATGTCTATG-ATATTCGTGTGGGT 495

```

```

QY 3057 taatctcataaacttagtctataaacagcttggtcctcactctgatgtctgtcctcaatcc 3116
|||||
Db 494 TATCTCATTAACATCTAGTATAAACAGTGGCTTCACCTTCATGATGTCTGCTCAAAATCC 435
|||||
QY 3117 ttttctcttaagaggtgtttatttataaagaaaaaaatgtaaaatgatagataataaa 3176
|||||
Db 434 TTTTCTCTNTANAGATGTTTATTATTAAG-AAAAAAATGTAATAATGATAGATAATAAA 376
|||||
QY 3177 agcttactaggttcttaaaagatgaactatccattatcattcagtaaatgaataattagtc 3236
|||||
Db 375 AGCCTTACTAGGTTCTTAAAGATGAACATATCCATATTTTCAGTAGATGAATAATTAGTCC 316
|||||
QY 3237 ttctctttgggcaccttggaaacagattcattcagatagtggtgggaatgtacatgtat 3296
|||||
Db 315 TTCTCTTTTGGCACCTTGGACAGATTTCATTCAGATAGTGGTGGAAATGTACATGTAT 256
|||||
QY 3297 ggttaagcattgtggcctagctactgaaaaatgaactcttatttttgcaggtg 3356
|||||
Db 255 GGTAAAGCATTGCTGGCTTAGTCACCTGAAAAATGTAGACTCTTATTTTGTATTCAGGTGG 196
|||||
QY 3357 aagttaagccatatgtcttgatgatcagctgtgtgatgaatgtcagggggcccgctgtg 3416
|||||
Db 195 AAGTTAAGCCATATGCTTGGATGATCAGCTGTGTGATGAATGTCAGGGGGCCCGTTGTG 136
|||||
QY 3417 gggggaaattgtctccattttctgtgctaatttacctgtctgcagtatctactgtgaat 3476
|||||
Db 135 GGGGAAATTTGCTCCATTTNTCTGAGAAAATGTTACATGTCAGCAGTATTCTGTGAAGT 76
|||||
QY 3477 attgctgggtgtctatccattctgctgctg 3506
|||||
Db 75 ATTGCTGGGTGCTATCCATTCATTCGTGCGG 46
|||||

```

## RESULT 11

```

BF903366/c 486 bp mRNA linear EST 18-JAN-2001
LOCUS QV1-MT0228-181200-583-g06 MT0228 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF903366
ACCESSION BF903366
VERSION BF903366.1 GI:12294825
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 486)
AUTHORS

```

```

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV1st2-QV1-MT0228-
181200-583-g06st3-2000-12-18st4-1)
Seq primer: puc 18 forward
High quality sequence stop: 484.
Location/Qualifiers
l..486
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

## FEATURES

## source

```

/clone_lib="WT0228"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
131 a 129 C 97 g 129 t
BASE COUNT
ORIGIN

```

Query Match	9.83;	Score 420.2;	DB 10;	Length 486;
Best Local Similarity	97.18;	Pred. No. 4.6e+43;		
Matches 428;	Conservative 0;	Mismatches 13;	Indels 0;	Gaps 0;
QY 292	taaaggctatgcattccctgcgtgtttc	caagatgaaagctctgtgcagcgtctcatgatgc	351	
DB				
486	TAAAGGCTATGCATTCCTGCTGT	TTCAGATGAAGCTCTGTGCAGGCTCTCATTGATGC	427	
QY 352	atgcattgaagaagatgaaacctacc	ctcttgtgtatcaagtcaccactatcaagataa	411	
DB				
426	ATGCATTGAAGAAGATGAAAACCT	CTACCTTTGTGTATCAAGTCCCACTATCAAGGATAA	367	
QY 412	gccagtcacagattcggccttggaa	cttcagtcacagtcgacttctgtatggatgggttcaca	471	
DB				
366	GCCAGTCCAGATTCCGCCCTTGG	AATCTCAGTCACAGTGACTTTGTGATGGATGGTTTACA	307	
QY 472	gccacttgaccacgaaanaactata	tatttggtyggtygttcctcgaccattacgagctgt	531	
DB				
306	GCACCTTGACCCACGAAAACATA	TATTTTGGTGGTGTTCCTCGACCAATTACGAGCTGT	247	
QY 532	ggagcttgcgattgcaatggatcgg	ctatcacgaggtgtgtcagcgtggattgatac	591	
DB				
246	GGATCTTGGATGATAATGGATCG	CTATATGGAGGTGTGTACGCTGGGATTTGATAC	187	
QY 592	cgaacctgagctaaatacccaaga	agctggagagttgccttctctaatcaacagag	651	
DB				
186	CGACCCCTGAGCTAAATAATACC	CAAAAGAGCTGGGAGAGTTGCGTTCTCTTAATCAACAGAG	127	
QY 652	ttacatagctgtatcagtcgccctt	gttcttcagctgcagcatggagatagataaacg	711	
DB				
126	TTACATAGCTGCTATCAGTGC	CCGCCCTTTGTTACGCTGCAGCATGGAGAGATAGATAAACG	67	
QY 712	ggtaagccttatactacattt	732		
DB				
66	GGTGAAGTTAAGCCATATGT	46		

RESULT	12
AW237641/c	
LOCUS	xm74h08.xl NCL_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2689983 3', EST 13-DEC-1999 linear mRNA 425 bp
DEFINITION	mRNA sequence.
ACCESSION	AW237641
VERSION	AW237641.1
KEYWORDS	GI:6570030
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 425)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-x@mail.nih.gov
COMMENT	Tissue Procurement: Christopher Moskaluk, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D.

156 a	58 c	145 t
Fatima Bonaldo. "		
1500552-1502855). Subtraction by Bento Soares and M.		
from a pool of 5,000 clones made from the same library		
(cloned) 1322376-1323911, 1456007-1456775, and		
hybridization reaction. The driver was PCR-amplified cDNAs		
prepared, and ss circles were made in vitro. Following HAP		
plasmid DNA from the normalized library NCI_CGAP_Kid3 was		
a modified polylinker: Site.1: Not I; Site.2: Eco RI;		
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with		
/lab_host="DH10B"		
/clone_lib="NCI_CGAP_Kid11"		
/clone="IMAGE:2689983"		
/db_xref="taxon:9606"		
/organism="Homo sapiens"		
1. .425		
Location/Qualifiers		
High quality sequence stop: 352.		
Seq primer: -400p from Gibco		
www-bio.lnl.gov/bbrp/image/image.html		
found through the I.M.A.G.E. Consortium/LLNL at:		
Clone distribution: NCI-CGAP clone distribution information can be		
DNA Sequencing by: Washington University Genome Sequencing Center		

Query Match	9.6%	Score 410.6;	DB 9;	Length 425;
Best Local Similarity	97.9%;	Pred. No. 7.6e-42;		
Matches 416;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0;
Qy	3845	aactctcgaaacatttcttagaggagagagagaataattaaaagagaaatgaacaat	3904	
Db	425	AAC TTC TCG CAAC AT TT TCT TAG AGG AG AG AG AAT AT TT AA AGA GAA AT TTA AGA GAA AT TTA AGA CAAT	366	
Qy	3905	agagcattttgggttttaataataattattgttaataataataacatataagaatacactttt	3964	
Db	365	AGAGTATTTTGGGTTTTAAATAAATTTATTTGTTAAATAATATAACAATAAGAAATACATTTT	306	
Qy	3965	atataataaccatgcacaataaacactatcggctctatctgcagcttttccccccaggga	4024	
Db	305	ATTAAATAACCATGCACAATAACACTATCGGTCTATCTGCAGAGTTTTCCTCCAGGGA	246	
Qy	4025	agtccttttgccttttcocttttcttttttttttttttttttttttttttggctctctcttt	4084	
Db	245	AGTCTTTTGGCTTTTCCTTTCTTTTTTTTTTTTTTTTTTTTTTTTCATCTTTTGTCCTCTCTTT	186	
Qy	4085	ttccatcccccttttaatttttaacagaaatggaggaagttaacaatttttaagggaa	4144	
Db	185	TTTCCATCCCTTTTAAATTTTTTTTAAACAGCANTGGAGGAAGTTTAAACAATTTTTTAATGGAA	126	
Qy	4145	agagcatgtttagagcaacaataatgcataagaacagtgcagcattataattaatttc	4204	
Db	125	AGAGCATGTTTAGAGCAACAATAATGCATAAAGCAAGACTTGAGCAGCATTTATTAATTAATTTTC	66	
Qy	4205	agggttttgagctgaacataatttctatttccctcaaaaagttaccacacatcagaaa	4264	
Db	65	AGGTTTGGAGCTGGAACATAATTTCAATATTCCTCAAAAGTTTTACCACCACATCAGAAA	6	
Qy	4265	aaaaa 4269		
Db	5	AAAAA 1		

RESULT	13
AI962924/c	
LOCUS	
DEFINITION	AI962924 wL2AH06.x1 NCI_CGAP_Utl1 Homo sapiens cDNA clone IMAGE:2508443 3', mRNA linear EST 09-MAR-2000
ACCESSION	AI962924
VERSION	AI962924.1
KEYWORDS	GI:5755637
SOURCE	EST. human.



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 433)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: crapsb@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 752 Std Error: 0.00  
Seq primer: -40up from Gibco  
High quality sequence stop: 362.

FEATURES  
source  
1. .433  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2508443"  
/clone\_lib="NCI CGAP Ut1"  
/tissue\_type="well-differentiated endometrial  
adenocarcinoma, 7 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.75 kb. Life Technologies catalog #:  
11538-014"  
153 a 59 c 68 g 151 t 2 others

Query Match 9.6%; Score 408.4; DB 9; Length 433;  
Best Local Similarity 97.9%; Pred. No. 1.4e-41;  
Matches 423; Conservative 0; Mismatches 8; Indels 1; Gaps 1;  
Qy 3845 aacttcctgcaacattttcttagagagagagagaaattattataaagaataaagaataa 3904  
Db 433 AACTTCCTGCAACATTTCTTAGAGGAGAGAGAGAAATATTAAAGAGAAATGAACAAT 374  
Qy 3905 agagatttttggttttaatttaattattgtttaataataataaataaataactttt 3964  
Db 373 AGAGTATTTTGGGTTTTAAATTAATTTATTTGTTAATAATAAACAATAAAGAAATCTTTT 314  
Qy 3965 attaaataaccatgcaacaataacactatcggtctatctgacagttttccccaggga 4024  
Db 313 ATTAAATTAACCATGCAACAAATACACTATCGGTCTATTGACAGTTTTCCTCCAGGGA 254  
Qy 4025 agtgccttttgcttttccctttcttttttttttttttttttttttttttttttttttt 4084  
Db 253 AGTGCCTTTTGCTTTTCCCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 195  
Qy 4085 ttctcatcccttttaatt 4144  
Db 194 TTTCCATCCCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 135  
Qy 4145 agagcattgttagagcaacaataatgataagcaagactgagcagcattataatttttc 4204  
Db 134 AGAGCATGTTAGAGCAACAATGCAATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 75  
Qy 4205 aggggttttgaggtgaaacataatttcattatccctcaaaaaagttaccacacacagaaa 4264  
Db 74 AGGGTTTGGAGCTGAACATAATTTTCAATTCCTCAAAAAAGTTTCCCCCCCATCAGAAA 15  
Qy 4265 aaaaaaataaaa 4276  
Db 14 AAAAAAAAAA 3

RESULT 14  
H71226/c  
LOCUS  
DEFINITION  
H71226  
IMAGE:214600 3', mRNA sequence.  
H71226  
H71226.1 GI:1043042  
EST.  
KEYWORDS  
SOURCE  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 438)  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins,  
M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore,  
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,  
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,  
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1184  
High quality sequence stops: 372  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1184 Std Error: 0.00  
Seq primer: Promega -2lml3  
High quality sequence stop: 372.

FEATURES  
source  
1. .438  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:3779504"  
/db\_xref="taxon:9606"  
/clone="IMAGE:214600"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAGAAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 135 a 90 c 98 g 112 t 3 others  
ORIGIN  
Query Match 9.3%; Score 399.6; DB 10; Length 438;  
Best Local Similarity 96.1%; Pred. No. 1.7e-40;  
Matches 419; Conservative 0; Mismatches 16; Indels 1; Gaps 1;  
Qy 3444 ctaatgttaactgtctgacttactgtg-aatatgtgggtgctgtatccatctcgt 3502  
Db 436 CTAAGTTTACCTGCTGCTGCAATATTACTGTAATATTCTGGGCTGTACCAATCTCGT 377  
Qy 3503 gctggcagggaattcccaagccccctggtgaaggaagcggtgaccgcctcgcatatt 3562  
Db 376 GCTGGCAGGGAATTCCCAACAGCCCGGTGAAGGNANAGGTGAACCCCTCGGCATATT 317

```
Qy 3563 tcattccgtggaactaaagataaactgcagtgctcatttccaggccctcagataaagtc 3622
Db 316 TCATTCCGCTGGAACCTAAAGGATAAAGTGCAGTGCCTCATTTTCAGGCCTCAGATAAGTGC 257
Qy 3623 actctctgttcattctgacccctctcctcaacctctctcaagctggaatgcctctttgtag 3682
Db 256 ACTCTCTGTTTCATCTTGACCCCTTCCTCAACCTCTTTCAGCGTGGCATGTCCTTTGTAG 197
Qy 3683 cagctgttaacttaactatagataatgaagaagaatgacctataataatagtggtttttagta 3742
Db 196 CAGTCTGTAACCTAATAGTATATAGTATAATGAAGAAGAACCTATATATAGTGTTTTGTGA 137
Qy 3743 gattctgtgtcactgcaacaataatgaacctcttttctgattgcatcggtgttcgcatg 3802
Db 136 GATTCTTGTGTCTACTGCAACCAATATGAACCTCTTTTCGTAATGGCATCGGGTTGCAUG 77
Qy 3803 gaagttttatctctgttctgttgctggaacccaagagagatccaacctctcctgcaaacattt 3862
Db 76 GAAGTTTATCTCTGTTTGTGGAACCAAGAGGATCCAAACTTCCTGCAACATTTT 17
Qy 3863 cttagaggagagagag 3878
Db 16 CTTAGAGGAGAGAG 1

RESULT 15
AI802170
LOCUS
DEFINITION tx25b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2270571 3',
mRNA sequence.
ACCESSION AI802170
VERSION AI802170.1 GI:5367642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 395)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 462 Std Error: 0.00
Seq primer: -400P from Gibco.
Location/Qualifiers
1..395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2270571"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (ClonesIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 130 a 58 c 61 g 146 t
ORIGIN
```

```
Query Match 9.0%; Score 383.8; DB 9; Length 395;
Best Local Similarity 98.2%; Pred. NO. 1.6e-38;
Matches 388; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3827 ggaacccaagagatcccaaaccttcctgcaacattttcttagagagagagagaaatatta 3886
Db 1 GGAACCAAGAGAGATCCCAACCTTCCTGCAACATTTCTTAGAGGAGAGAGAAATATTA 60
Qy 3887 aaagagaatgaacaatagagattttgggtttttaataattattgttaataatata 3946
Db 61 AAAGAGAAATGAACAATAGAGTATTTGGGTTTAAATTAATTAATTTGTTAAATAATATA 120
Qy 3947 acatataagaataactttttattataaataaaccatgcaacaataaacactatcggtctatctga 4006
Db 121 ACATATAAGAATACACTTTTATTAAAAATAACCATGCAACAATAACACTATCGGTCTATCTGA 180
Qy 4007 cagtttttccccccaggggaagtgcttttgcctttctcttcttcttttttttttttttctatct 4066
Db 181 CAGTTTTTCCCCCAGGGAAGTGCCTTTTCCATCCCTTTTAAATTTTCTTTCTTTTCTTTTTCATC 240
Qy 4067 ttttggctctctctcttttttccatcccttttttaattttttaacagcaatggaggaagt 4126
Db 241 TTTTGTGCTCTCTCTCTTTTTCATCCCTTTTAAATTTTCTTTCTTTTCTTTTCTTTTCTTTTCT 300
Qy 4127 taacaatttttaaggaagagcatgttagagcaacaacaatgcataagcaagactgagca 4186
Db 301 TAACAATTTTAAUGGAAGAGCATGTTAGAGCAACAATAATGCAATAAGCAAGACTGAGCA 360
Qy 4187 gcattataatttaattttcagggtttttgaggctgaa 4221
Db 361 GCATTATAATTAATTTTCAGGGTTTTCAGGGCTGAA 395
```

Search completed: September 1, 2002, 09:22:57  
Job time: 11116 sec

---



OM of: US-09-730-559B-45 to: A\_Geneseq\_032802:\* out\_format : pfs  
Date: Sep 8, 2002 8:09 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters: -MODEL=frame+np.model -DEV=xlh  
-Q=/cgn2\_1/USPFO.spool/US09730559/runat\_03092002\_150444\_3252/app\_query.fasta\_1.86  
-DB=A\_Geneseq\_032802 -QPMF=Fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09730559 -CGNL\_1.0 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-730-559B-45  
Query length: 26  
Database: A\_Geneseq\_032802:\*  
Database sequences: 747574  
Database length: 11107396  
Search time (sec): 79.760000

score_list:	Strd Orig	zScore	EScore	Len	! Documentation
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAE12174 +	36.00	142.31	9.40	95	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAE12174 +	36.00	134.75	10.37	22	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAE12174 +	35.00	126.66	18.46	36	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU61249 +	35.00	119.86	20.16	78	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABE68264 +	34.00	132.47	27.43	11	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW27851 +	33.00	128.84	46.07	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW11794 +	33.00	127.52	46.87	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW11769 +	33.00	117.78	53.18	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE39625 -	33.00	117.65	53.28	39	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW14111 +	33.00	117.48	53.43	39	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE67128 -	33.00	117.48	53.43	40	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE3837 -	33.00	116.86	53.83	43	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW16678 +	33.00	114.83	55.26	54	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE10933 +	33.00	114.06	55.82	54	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE1340 +	33.00	113.57	56.17	63	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AAE38611 +	33.00	109.23	59.43	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE2881 +	33.00	107.04	61.15	13	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE41178 +	32.00	129.10	73.57	66	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE36095 +	32.00	122.90	79.75	13	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE81717 +	32.00	118.86	84.04	21	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE36342 -	32.00	116.06	87.15	29	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE6538 +	32.00	115.55	87.73	31	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE2962 +	32.00	115.52	87.77	31	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE01071 -	32.00	111.05	93.01	52	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE73536 -	32.00	110.87	93.22	54	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE40603 +	32.00	109.73	94.62	61	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE4788 +	32.00	109.71	94.63	61	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE40602 +	32.00	109.70	94.65	61	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE24787 +	32.00	109.46	94.94	63	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE40601 +	32.00	109.45	94.96	63	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE24786 +	32.00	109.41	95.01	63	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE24850 +	32.00	108.83	95.72	68	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE24849 +	32.00	108.64	95.96	69	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE24848 +	32.00	108.58	96.03	70	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE24847 +	32.00	105.07	100.52	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AAE82261 +	32.00	105.07	100.52	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU02494 +	32.00	105.07	100.52	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE2295 +	32.00	105.07	100.52	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU16364 +	31.00	120.43	131.96	11	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAE37429 -	31.00	119.61	133.36	12	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE10558 -	31.00	119.20	134.07	13	

/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE10557 - 31.00 119.13 134.19  
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAE10555 - 31.00 119.07 134.30  
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU40035 - 31.00 117.50 137.07  
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU34357 + 31.00 116.21 139.37  
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU37125 + 31.00 116.21 139.37  
seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAE12174

seq\_documentation\_block:  
ID AA12174 standard; Protein; 95 AA.  
XX  
AC AA12174;  
XX  
DT 18-JUN-1999 (first entry)  
XX  
XX  
DE Human 5' EST secreted protein SEQ ID NO: 487.  
XX  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
XX  
OS Homo sapiens.  
XX  
PN W090905554-A2.  
XX  
PD 11-FEB-1999.  
XX  
XX  
PF 31-JUL-1998; 98WO-IB01238.  
XX  
PR 01-AUG-1997; 97US-0905134.  
XX  
(BEST ) GENSET.  
XX  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
XX  
XX WPI: 1999-153784/13.  
XX  
XX N-PSDB: AAX41007.  
XX  
XX New nucleic acids encoding human secreted proteins - obtained from  
XX cDNA libraries prepared from kidney, fetal kidney, dystrophic  
XX muscle, muscle and heart tissue  
XX  
XX Claim 34; Page 566; 622pp; English.

AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for human secreted proteins, and encode the proteins given in AAY01602 and AAY11994 to AAY12260, respectively. The proteins given represent the signal peptide and an N-terminal fragment of a secreted protein. The nucleic acid sequences can be used for producing secreted human gene products. They can also be used to develop products for diagnosis and therapy. The proteins obtained may have cytokine activity, cell proliferation/differentiation activity, haematopoiesis regulating activity, tissue growth regulating activity, reproductive hormone regulating activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, tumour inhibition activity or other activities. The products can be used in forensic, gene therapy and chromosome mapping procedures. The sequences can also be used for obtaining corresponding promoter sequences. The nucleic acids encoding the signal peptide can be used for directing extracellular secretion of a polypeptide or the insertion of a polypeptide into a membrane, or importing a polypeptide into a cell.

Sequence 95 AA;

alignment\_scores:  
Quality: 36.00 Length: 7  
Ratio: 5.143 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x AAY12174 ..  
Align seg 1/1 to: AAY12174 from: 1 to: 95

1 GGGCTTAATATTATTCATAGA 21  
|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT: AAY31836

seq\_documentation\_block:

ID AAY31836 standard; Protein: 227 AA.

XX AC AAY31836;

DT 06-DEC-1999 (first entry)

XX Human foetal kidney secreted protein pm633\_1.

XX Secreted protein; pm633\_1; human; therapy; diagnosis; vaccine;  
XX kidney.

XX OS Homo sapiens.

XX Key Location/Qualifiers  
XX Peptide 28..40  
XX Protein /note= "signal peptide"

XX FT 41..227  
XX FT /note= "mature protein"

XX PN WO9947555-A1.

XX 23-SEP-1999.

XX 18-MAR-1999; 99WO-US05939.

XX 20-MAR-1998; 98US-0078803.

XX 17-MAR-1999; 99US-0078803.

XX (GEMY ) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ;

XX WPI; 1999-562059/47.

XX N-PSDB; AA219900.

XX New polynucleotides derived from murine foetal cell cDNA libraries,  
XX potentially used as, e.g. vaccines -

XX Claim 25(a); Page 101-102; 107pp; English.

XX This is the predicted amino acid sequence of a novel human secreted  
XX protein, pm633\_1, as deduced from an isolated foetal kidney cDNA  
XX clone (see AA219900). An alternative sequence for pm633\_1 (from  
XX ORF2) is given in AAY31837. The invention provides new human secreted  
XX proteins (see AAY31828-38) and polynucleotides (see AA219893-901)  
XX isolated from foetal cell, adult blood, adult brain and foetal  
XX kidney cDNA libraries. They are predicted to have biological  
XX activities which would make them suitable for treating, preventing  
XX or ameliorating medical conditions in humans and animals, although  
XX no supporting data are given. Suggested activities include  
XX nutritional, cytokine, tissue growth, cell proliferation and  
XX differentiation, immunostimulant (e.g. as vaccine),  
XX immunosuppressive, haematopoiesis regulating, activin or inhibin,  
XX chemotactic or chemokinetic, haemostatic or thrombolytic,  
XX receptor/ligand activity, antiinflammatory, cadherin or tumour  
XX invasion suppressor, and tumour inhibition activities.

XX Sequence 227 AA;

alignment\_scores:  
Quality: 36.00 Length: 7  
Ratio: 5.143 Gaps: 0  
Percent similarity: 100.000 Percent Identity: 85.714

alignment\_block:  
US-09-730-559B-45 x AAY31836 ..

Align seg 1/1 to: AAY31836 from: 1 to: 227

1 GGGCTTAATATTATTCATAGA 21  
|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT: AAU61249

seq\_documentation\_block:

ID AAU61249 standard; Protein: 360 AA.

XX AC AAU61249;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #22145.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX DR N-PSDB; AAS59615.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
XX vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris -

XX Example 1; SEQ ID No 22444; 1069pp; English.

XX Sequences AA039105-AA068017 represent Propionibacterium acnes immunogenic  
XX polypeptides. The proteins and their associated DNA sequences are used in  
XX the treatment, prevention and diagnosis of medical conditions caused by  
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
XX P. acnes is also involved in infections of bone, joints and the central  
XX nervous system, however it is particularly involved in the inflammatory  
XX lesions associated with acne vulgaris. A method for detecting the  
XX presence or absence of P. acnes in a patient comprises contacting a  
XX sample with a binding agent that binds to the proteins of the invention  
XX and determining the amount of bound protein in the sample. The  
XX polypeptides may be used as antigens in the production of antibodies  
XX specific for P. acnes proteins. These antibodies can be used to  
XX downregulate expression and activity of P. acnes polypeptides and  
XX therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 360 AA;

alignment\_scores:  
Quality: 35.00 Length: 7  
Ratio: 5.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 71.429

alignment\_block:  
US-09-730-559B-45 x AAU61249 ..

Align seg 1/1 to: AAU61249 from: 1 to: 360

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||  
30 GlyLeuAsnValValHisArg 36

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT:ABB68264

seq\_documentation\_block:

ID ABB68264 standard; Protein; 788 AA.

XX AC ABB68264;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 31584.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB; ABL12367.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX Disclosure; SEQ ID NO 31584; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (AB57737-AB57207).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 788 AA;  
alignment\_scores:  
Quality: 35.00 Length: 8  
Ratio: 4.375 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45 x ABB68264 ..

Align seg 1/1 to: ABB68264 from: 1 to: 788

1 GGGCTTAATATTATTCATAGATCG 24  
|||||:|||||:|||||  
118 GlyLeuAsnIleIleHisLysAla 125

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT:AAW27851

seq\_documentation\_block:

ID AAW27851 standard; Protein; 115 AA.

XX AC AAW27851;

XX 21-JUL-1998 (first entry)

XX Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
KW toxic shock syndrome.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Misc-difference 1..115

FT /note= "residues designated X are not defined in  
FT the specification"

FT Misc-difference 1

FT /note= "Met encoded by CTG"

XX WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI: 1997-424969/39.

DR N-PSDB; AAW83814.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
PT to isolate antimicrobial compounds, and in vaccines against S.  
PT aureus infection

XX Claim 6; Page 310; 989pp; English.

XX The present sequence represents a Staphylococcus aureus protein of  
CC unknown function. The DNA sequence was isolated from a library of  
CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can  
CC be used in the construction of ribozymes and antisense sequences to  
CC control the expression of Staphylococcal genes. The DNA sequence is  
CC also useful as a source of regulatory elements for the control of  
CC bacterial gene expression. The present protein may be used to produce  
CC vaccines to enable a host to produce specific antibodies with  
CC antibacterial action. These vaccines and antibodies would protect

CC a host against invasion by *S. aureus*, and conditions relating to  
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
CC skin syndrome, and toxic shock syndrome.  
XX  
SQ Sequence 115 AA;

alignment\_scores:  
Quality: 34.00 Length: 8  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 87.500 Percent Identity: 87.500

alignment\_block:  
US-09-730-559B-45 x AAW27851 ..

Align seg 1/1 to: AAW27851 from: 1 to: 115

2 GGCTTAATATTATTCATAGATCGA 25  
||||| |||||||||  
12 GlyLeuTyrLeuPheIleAspArg 19

seq\_name: /SIDSl/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW11794

seq\_documentation\_block:  
ID AAW11794 standard; Protein; 109 AA.

AC AAW11794;

XX 24-APR-1997 (first entry)

DE Early onset Alzheimer's disease gene product (uterine cancer).

XX Early onset Alzheimer's disease; EOAD; diagnosis; therapy;

KW expressed sequence tag; EST; uterus cancer.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 86 /note= "amino acid residue at position 86 is  
FT unidentified owing to degeneracy of the  
FT nucleotide sequence"

FT Misc-difference 89 /note= "amino acid residue at position 89 is  
FT unidentified owing to degeneracy of the  
FT nucleotide sequence"

FT Misc-difference 97 /note= "amino acid residue at position 97 is  
FT unidentified owing to degeneracy of the  
FT nucleotide sequence"

FT Misc-difference 105 /note= "amino acid residue at position 105 is  
FT unidentified owing to degeneracy of the  
FT nucleotide sequence"

XX W09701573-A2.

XX 16-JAN-1997.

XX 26-JUN-1996; 96WO-US11002.

XX 28-JUN-1995; 95US-0000590.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

XX Karran EH;

XX WPI; 1997-100161/09.

DR N-PSDB; AAT59474.

XX New Early Onset Alzheimer's Disease gene - used to develop prods.  
PT for diagnosis, detection of pre-disposition to, or treatment of

PT Alzheimer's disease  
XX  
PS Claim 3; Page 75; 97pp; English.  
XX  
CC A set of polypeptides (AAW11792-97) have sequences deduced from the  
CC 3 reading frames, of both strands, of a human uterine cancer cDNA  
CC clone (AAT59474) that corresponds to an expressed sequence tag (EST)  
CC of the early onset Alzheimer's disease (EOAD) gene. EOAD gene  
CC products (see also AAW11768-91) can be expressed in host cells and  
CC used to screen for agonists or antagonists useful in EOAD therapy,  
CC or to raise antibodies useful in the diagnosis of EOAD or  
CC predisposition to EOAD.  
CC (N.B. in-frame stop codons in the 6 reading frames of the cDNA  
CC clone are ignored in the translated polypeptide sequences given in  
CC the specification)  
XX  
SQ Sequence 109 AA;

alignment\_scores:  
Quality: 33.00 Length: 7  
Ratio: 4.714 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:  
US-09-730-559B-45 x AAW11794 ..

Align seg 1/1 to: AAW11794 from: 1 to: 109

4 CTTAATATTATTCATAGATCG 24  
||||| |||||||  
34 LeuAsnIleLeuHisArgSer 40

seq\_name: /SIDSl/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW11769

seq\_documentation\_block:

ID AAW11769 standard; Protein; 127 AA.

XX AC AAW11769;

XX 24-APR-1997 (first entry)

DE Early onset Alzheimer's disease gene product (amygdala).

XX Early onset Alzheimer's disease; EOAD; diagnosis; therapy;  
KW expressed sequence tag; EST; amygdala.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 3 /note= "amino acid residue at position 3 is  
FT unidentified owing to degeneracy of the  
FT nucleotide sequence"

FT Misc-difference 46 /note= "amino acid residue at position 46 is  
FT unidentified owing to degeneracy of the  
FT nucleotide sequence"

FT Misc-difference 103 /note= "amino acid residue at position 103 is  
FT unidentified owing to degeneracy of the  
FT nucleotide sequence"

XX W09701573-A2.

XX 16-JAN-1997.

XX 26-JUN-1996; 96WO-US11002.

XX 28-JUN-1995; 95US-0000590.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.



XX Karan EH;  
 XX WPI: 1997-100161/09.  
 DR N-PSDB; AAT59470.  
 XX  
 XX New Early Onset Alzheimer's Disease gene - used to develop prods.  
 PT for diagnosis, detection of pre-disposition to, or treatment of  
 PT Alzheimer's disease  
 XX  
 XX Claim 3; Page 55; 97pp; English.  
 XX  
 CC A set of polypeptides (AAW11768-73) have sequences deduced from the  
 CC 3 reading frames, of both strands, of a human amygdala cDNA clone  
 CC (AAT59470) that corresponds to an expressed sequence tag (EST) of the  
 CC early onset Alzheimer's disease (EOAD) gene. EOAD gene products  
 CC (see also AAW11774-97) can be expressed in host cells and used to  
 CC screen for agonists or antagonists useful in EOAD therapy, or to  
 CC raise antibodies useful in the diagnosis of EOAD or predisposition  
 CC to EOAD.  
 CC (N.B. in-frame stop codons in the 6 reading frames of the cDNA  
 CC clone are ignored in the translated polypeptide sequences given in  
 CC the specification)  
 XX  
 XX Sequence 127 AA;  
 SQ  
  
 alignment\_scores:  
 Quality: 33.00 Length: 7  
 Ratio: 4.714 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 85.714  
  
 alignment\_block:  
 US-09-730-559B-45 x AAW11769 ..  
 Align seg 1/1 to: AAW11769 from: 1 to: 127  
  
 4 CTTAATATTATTCATGATCG-24  
 |||||.....  
 25 LeuAsnIleLeuHisArgSer 31  
  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW39625  
  
 seq\_documentation\_block:  
 ID AAW39625 standard; Protein: 390 AA.  
 XX  
 AC AAW39625;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 XX Human polypeptide SEQ ID NO 2770.  
 DE  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 XX 26-JUL-2001.  
 PD  
 XX 26-DEC-2000; 2000WO-US34263.  
 PF  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-052317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AAI58781.  
 DR  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PT  
 XX Example 4; SEQ ID NO 2770; 10078pp; English.  
 PS  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous system diseases, such as  
 CC localised neuropathies and central nervous system diseases, amyotrophic  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 XX Sequence 390 AA;  
 SQ  
  
 alignment\_scores:  
 Quality: 33.00 Length: 6  
 Ratio: 5.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
  
 alignment\_block:  
 US-09-730-559B-45/rev x AAM39625 ..  
 Align seg 1/1 to: AAM39625 from: 1 to: 390  
  
 18 ATGATATATTAAGCCC 1  
 |||||.....  
 35 MetAsnAsnIleLysPro 40  
  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW41411  
  
 seq\_documentation\_block:  
 ID AAW41411 standard; Protein: 396 AA.  
 XX  
 AC AAW41411;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 XX Human polypeptide SEQ ID NO 6342.  
 DE  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 XX 26-JUL-2001.  
 PD  
 XX 26-DEC-2000; 2000WO-US34263.  
 PF  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-052317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.

PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AAI60567.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 2; SEQ ID NO 6342; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 396 AA;

alignment\_scores:  
 Quality: 33.00 Length: 6  
 Ratio: 5.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-730-559B-45/rev x AAM41411 ..  
 Align seg 1/1 to: AAM41411 from: 1 to: 396  
 18 ATGAATATATATAGCCC 1  
 41 MetAsnAsnIleIlyPro 46

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAG67128  
 seq\_documentation\_block:  
 ID AAG67128 standard; Protein; 404 AA.  
 XX  
 AC AAG67128;  
 XX  
 DT 13-NOV-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human enzyme.  
 XX  
 KW Human; enzyme; cancer; neurological disorder; epilepsy; stroke;

KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;  
 KW multiple sclerosis; Parkinson's disease; amyotrophic lateral sclerosis;  
 KW meningitis; schizophrenic disorder; neuroskeletal disorder; allergy;  
 KW Addison's disease; autoimmune disease; anemia; asthma; Crohn's disease;  
 KW adult respiratory distress syndrome; atopic dermatitis; psoriasis;  
 KW diabetes mellitus; osteoporosis; pancreatitis; rheumatoid arthritis;  
 KW infection; genetic disorder; muscular dystrophy; Gaucher's disease;  
 KW Huntington's chorea; sickle cell anemia; thalassemia; atherosclerosis;  
 KW Von Willebrand's disease; Wilms' tumour; cell proliferative disorder;  
 KW leukemia; hepatitis; cirrhosis; arteriosclerosis; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Modified-site 136  
 FT /note= "potential phosphorylation site"  
 FT Modified-site 155  
 FT /note= "potential phosphorylation site"  
 FT Modified-site 192  
 FT /note= "potential phosphorylation site"  
 FT Modified-site 293  
 FT /note= "potential phosphorylation site"  
 FT Modified-site 301  
 FT /note= "potential phosphorylation site"  
 FT  
 XX WO200164896-A2.  
 XX  
 XX 07-SEP-2001.  
 XX  
 XX 01-MAR-2001; 2001WO-US06806.  
 XX  
 XX 01-MAR-2000; 2000US-0186307.  
 XX 28-MAR-2000; 2000US-0192532.  
 XX 30-MAR-2000; 2000US-0193578.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Tang YT, Lu DAM, Bandman O, Yue H, Azimzai Y, Lal P, Burford N;  
 PI Baughn MR;  
 PI  
 DR WPI: 2001-550184/61.  
 DR N-PSDB; AAM75156.  
 XX  
 XX Novel human enzyme molecule useful for treating and preventing, e.g.,  
 PT cancer, genetic disorders, neurological disorders, autoimmune and  
 PT inflammatory disorders -  
 PS  
 PS Claim 1; Page 118; 154pp; English.  
 XX  
 XX The present sequence represents a human enzyme. The enzyme polynucleotide  
 CC and polypeptide are useful for diagnosis, treatment and prevention of  
 CC cancers, neurological disorders (e.g. epilepsy, stroke, Alzheimer's  
 CC disease, Pick's disease, Huntington's disease, dementia, multiple  
 CC sclerosis, Parkinson's disease, amyotrophic lateral sclerosis, bacterial  
 CC and viral meningitis, schizophrenic disorders and neuroskeletal  
 CC disorders), autoimmune/inflammatory disorders (e.g. allergies,  
 CC Addison's disease, autoimmune diseases, adult respiratory distress  
 CC syndrome, anemia, asthma, Crohn's disease, atopic dermatitis, diabetes  
 CC mellitus, osteoporosis, pancreatitis, psoriasis, rheumatoid arthritis,  
 CC and viral, bacterial, fungal, parasitic, protozoal and helminthic  
 CC infections), genetic disorder (e.g. Duchenne and Becker muscular  
 CC dystrophy, Gaucher's disease, Huntington's chorea, sickle cell anemia,  
 CC thalassemia, Von Willebrand's disease and Wilms' tumour), and cell  
 CC proliferative disorder (e.g. atherosclerosis, leukemia, hepatitis,  
 CC cirrhosis, and arteriosclerosis). The polynucleotide is also useful in  
 XX somatic or germline gene therapy.  
 XX  
 SQ Sequence 404 AA;

alignment\_scores:  
 Quality: 33.00 Length: 6  
 Ratio: 5.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-730-559b-45/rev x AAG67128 ..

Align seg 1/1 to: AAG67128 from: 1 to: 404

18 ATGAATAATATTAAAGCCC 1

|||||

1 MetAsnAsnIleLysPro 6

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT: AAB93837

## seq\_documentation\_block:

ID AAB93837 standard; Protein: 434 AA.

XX AAB93837;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:13667.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

XX EPI074617-A2.

PN 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

PS Claim 8; SEQ ID 13667; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13632 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 434 AA;

## alignment\_scores:

Quality: 33.00 Length: 6

Ratio: 5.500 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-730-559b-45/rev x AAB93837 ..

Align seg 1/1 to: AAB93837 from: 1 to: 434

18 ATGAATAATATTAAAGCCC 1

|||||

42 MetAsnAsnIleLysPro 47

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT: AAW16678

## seq\_documentation\_block:

ID AAW16678 standard; Protein: 548 AA.

XX AAW16678;

DT 20-AUG-1997 (first entry)

DE Lawsonia intracellularis GroEL.

KW Intestinal disease; porcine proliferative enteropathy; vaccine;  
KW GroEL; immunotherapy; antibody; diagnosis; heat shock protein.

OS Lawsonia intracellularis.

XX WO9720050-A1.

PN 05-JUN-1997.

XX 29-NOV-1996; 96WO-AU00767.

XX 30-NOV-1995; 95AU-0006911.

XX 30-NOV-1995; 95AU-0006910.

XX (DARA-) DARATECH PTY LTD.

XX (PIGR-) PIG RES & DEV CORP.

XX Hasse D, Panaccio M;

XX WPI; 1997-310605/28.

XX N-PSDB; AAT69201.

XX Vaccine for treating or preventing Lawsonia intracellularis

XX infection - especially in pigs, containing non-pathogenic form of

XX bacterium or its components

XX Claim 10; Page 38-42; 94pp; English.

XX The GroEL heat shock protein (AAW16678) of Lawsonia intracellularis

XX can be used in vaccines to protect birds and animals against

XX intestinal diseases, esp. to protect pigs against porcine

XX proliferative enteropathy. It is the expression product of a DNA

XX molecule (AAH69201) obtd. by screening an L. intracellularis library

XX with rabbit anti-L. intracellularis sera. GroEL, GroES (AAW16679)

XX and other L. intracellularis polypeptides (AAW16680-85) can be

XX administered as recombinant polypeptides or expressed as

XX recombinant vaccines utilising bacterial, fungal or viral vectors.

XX Antibodies raised against the polypeptides may be useful in

XX immunotherapy, diagnosis of infection and detection.

XX Sequence 548 AA;

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:  
US-09-730-559B-45 x AAW16678 ..

Align seg 1/1 to: AAW16678 from: 1 to: 548

1 GGGCTTAATATATTCATAGATCG 24  
|||||  
439 GlyLeuAsnIleIleArgSer 446

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA10933

seq\_documentation\_block:

ID AAB10933 standard; protein; 599 AA.

AC AAB10933;

DT 07-FEB-2001 (first entry)

DE S. pombe rrn3p protein fragment.

KW RNA polymerase I transcription factor TIF-1A; rrn3p; antitumor; cancer;  
KW antiproliferative; cell proliferation; treatment; tissue regeneration.

OS Schizosaccharomyces pombe.

PN WO200055316-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-DE00767.

PR 17-MAR-1999; 99DE-1011992.

PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

PI Grummt I, Vingron M;

DR WPI; 2000-587527/55.

PT New DNA encoding the transcription factor TIF-1A, useful for preventing  
PT or treating diseases associated with abnormal cell proliferation,  
PT particularly tumors

PS Example 1; Fig 1; 38pp; German.

XX This invention describes a novel DNA sequence (I) that encodes the RNA  
XX polymerase I transcription factor TIF-1A which has antitumor,  
XX antiproliferative and proliferation-inducing activity. The invention also  
XX describes (1) DNA (Ia) encoding a protein (II) with the biological  
XX activity of TIF-1A; (2) a ribozyme (R) corresponding to (I) or (Ia) and  
XX able to bind specifically to, and cleave, its transcribed RNA so as to  
XX reduce or inhibit synthesis of the corresponding protein; (3) an  
XX antisense RNA (AS) with binding properties similar to R; (4) an  
XX expression vector that contains (I) (Ia) or sequences that encode R or  
XX AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (II)  
XX encoded by (I) or (Ia); (7) preparation of TIF-1A or (II) by culturing  
XX cells of (6); (8) ligands that bind to TIF-1A or (II); (9) antagonists  
XX that weaken or block the activity of TIF-1A or (II); (10) a diagnostic  
XX method for detecting abnormal TIF-1A expression; and (11) kit for  
XX carrying out the method in (10). (I), and similar sequences that encode  
XX proteins with equivalent activity, expression vectors containing them, or  
XX their expression products are used to treat or prevent disorders  
XX associated with reduced cellular proliferation, to stimulate cellular  
XX proliferation, and to promote tissue regeneration, e.g. after injury or  
XX radiation therapy. Ribozymes, antisense sequences directed against (I),  
XX also ligands and antagonists of TIF-1A, are used to treat or prevent  
XX disorders associated with excessive cellular proliferation and to inhibit  
XX proliferation, especially in treatment of cancers. (I) and specific

CC ligands for TIF-1A (particularly antibodies (Ab)) are also useful for  
CC diagnosis of altered TIF-1A expression by (in)direct determination of the  
CC concentration, length and/or sequence of TIF-1A or its mRNA, e.g. for  
CC detecting mutations. Ab can also be used for immunoprecipitation of  
CC TIF-1A and for isolation of related sequences from cDNA expression  
CC libraries. (I) allows recombinant production of TIF-1A in sufficient  
CC quantities for therapeutic use.

SQ Sequence 599 AA;

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45 x AAB10933 ..

Align seg 1/1 to: AAB10933 from: 1 to: 599

1 GGGCTTAATATTCATAGATCG 24  
|||||  
488 GlyLeuGluIleLeuHisArgSer 495

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU51340

seq\_documentation\_block:

ID AAU51340 standard; Protein; 634 AA.

AC AAU51340;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #12236.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59551.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
XX vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris

PS Example 1; SEQ ID No 12535; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 634 AA;

alignment\_scores:                   Quality: 33.00                   Length: 7  
                                      Ratio: 4.714                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x AAU51340   ..

Align seg 1/1 to: AAU51340 from: 1 to: 634

2 GGCTTATATATTCATAGAT 22

|||||:::|||||  
58 GlyLeuValLeuPheIleAsp 64

seq\_name: /SDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: AAR58611

seq\_documentation\_block:

ID AAR58611 standard; Protein; 1045 AA.

AC AAR58611;

DT 28-APR-1995 (first entry)

DE Yeast HMG-CoA reductase 2.

KW HMG-CoA reductase 2; 3-hydroxy-3-methyl:glutaryl coenzyme A;

KW squalene; sterol.

XX Saccharomyces cerevisiae.

XX US5349126-A.

XX 20-SEP-1994.

XX 12-OCT-1990; 90US-0596467.

XX 12-OCT-1990; 90US-0596467.

PR 14-AUG-1992; 92US-0934374.

XX (STAD ) AMOCO CORP.

XX Chappell J, Saunders CA, Wolf FR;

XX WPI; 1994-302280/37.

DR N-PSDB; AAQ70611.

XX Transgenic plants with increased insect resistance - by  
PT transformation with DNA encoding HMG-CoA activity to increase  
PT squalene and sterol accumulation

XX Disclosure; Page 53; 58pp; English.

XX AAQ70611 codes for a yeast HMG-CoA reductase 2, AAR58611. Using the  
CC A. tumefaciens vector pKYLX71, AAQ70611 was used to produce transgenic  
CC plants with increased HMG-CoA reductase activity. Plants with

CC increased HMG-CoA reductase activity were found to have greater  
CC squalene and sterol accumulation, and therefore increased insect  
CC resistance.

XX Sequence 1045 AA;

alignment\_scores:

                                      Quality: 33.00                   Length: 7  
                                      Ratio: 4.714                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x AAR58611   ..

Align seg 1/1 to: AAR58611 from: 1 to: 1045

4 CTTAATATATTCATAGATCG 24

:::|||||  
360 IleAsnIleIleHisArgSer 366



```
Search information block:
Query: US-09-730-559B-45
Query length: 26
Database: Issued_Patents_AA:
Database sequences: 231628
Database length: 24425594
Search time (sec): 29.910000
```

```

/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-919-627-3 + 30.00 104
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-096-245-3 + 30.00 104
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-713-828-4 + 30.00 104
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-713-828-5 + 30.00 104
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-919-627-4 + 30.00 104
seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-596-467-6

seq_documentation_block:
; Sequence 6, Application US/07596467
; Patent No. 5306862
; GENERAL INFORMATION:
; APPLICANT: Chappell, J.
; APPLICANT: Saunders, Court A.
; APPLICANT: Cuellar, R.
; APPLICANT: Wolf, Fred R.
; TITLE OF INVENTION: Method and Composition for Increasing
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker & Milnamow
; STREET: 180 N. Stetson St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/596.467
; FILING DATE: 19901012
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 616-5400
; TELEFAX: 312 616-5460
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-596-467-6

alignment_scores:
Quality: 33.00 Length: 7
Ratio: 4.714 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x US-07-596-467-6 ..
Align seg 1/1 to: US-07-596-467-6 from: 1 to: 1045

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-934-374-6

seq_documentation_block:
; Sequence 6, Application US/07934374
; Patent No. 5349126
; GENERAL INFORMATION:
; APPLICANT: Chappell, J.
; APPLICANT: Saunders, Court A.
; APPLICANT: Cuellar, R.
; APPLICANT: Wolf, Fred R.

```

```
; TITLE OF INVENTION: Process and Composition for Increasing
; TITLE OF INVENTION: Sterol Accumulation in Higher Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Suker & Milnamow
; STREET: 180 N. Stetson St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/934,374
; FILING DATE: 19920814
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 616-5400
; TELEFAX: 312 616-5460
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-934-374-6

alignment_scores:
  Quality: 33.00      Length: 7
  Ratio: 4.714      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x US-07-934-374-6 ..
Align seg 1/1 to: US-07-934-374-6 from: 1 to: 1045

4 CTTAATATTATTCATGATCG 24
  ::::::::::::::::::::
360 IleAsnIleIleHisArgSer 366

seq_name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:US-07-783-861C-6

seq_documentation_block:
; Sequence 6, Application US/07783861C
; Patent No. 5460949
; GENERAL INFORMATION:
; APPLICANT: Saunders, Court A.
; APPLICANT: Wolf, Fred R.
; APPLICANT: Mukharji, Indrani
; TITLE OF INVENTION: A Method and Composition for Increasing
; TITLE OF INVENTION: the Accumulation of Squalene and Specific Sterols in
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept.
; STREET: 200 East Randolph St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680-0703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/783.861C
; FILING DATE: 19911028
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613.380
; FILING DATE: 15-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5460949vall B.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 856-7180
; TELEFAX: 312 856-4972
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1045 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-783-861C-6

alignment_scores:
  Quality: 33.00      Length: 7
  Ratio: 4.714      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x US-07-783-861C-6 ..
Align seg 1/1 to: US-07-783-861C-6 from: 1 to: 1045

4 CTTAATATTATTCATGATCG 24
  ::::::::::::::::::::
360 IleAsnIleIleHisArgSer 366

seq_name: /cgn2_6/ptodata/1/laa/6B_COMB.pep:US-09-433-994-2

seq_documentation_block:
; Sequence 2, Application US/09433994
; Patent No. 6326172
; GENERAL INFORMATION:
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Warren, Patrick V.
; APPLICANT: Ingraham, Karen I.
; APPLICANT: Chalker, Alison F.
; APPLICANT: So, Chi Young
; APPLICANT: Holmes, David J.
; APPLICANT: Van Horn, Stephanie
; TITLE OF INVENTION: ytgp
; FILE REFERENCE: GM10246
; CURRENT APPLICATION NUMBER: US/09/433,994
; CURRENT FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-433-994-2

alignment_scores:
  Quality: 32.00      Length: 8
  Ratio: 4.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 62.500

alignment_block:
US-09-730-559B-45/rev x US-09-433-994-2 ..
Align seg 1/1 to: US-09-433-994-2 from: 1 to: 540

24 CGATCTATGTAATATTAAGCCC 1
```





```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,380
; FILING DATE: 15-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5460949vall B.
; TELEPHONE: 312 856-7180
; TELEFAX: 312 856-4972
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1054 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-783-861C-2

alignment_scores:
    Quality: 32.00      Length: 7
    Ratio: 4.571       Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 71.429

alignment_block:
US-09-730-559B-45 x US-07-783-861C-2 ..

Align seg 1/1 to: US-07-783-861C-2 from: 1 to: 1054

4 CTTAATATTATTATGATCG 24
:::|||||:|||||
361 MetAsnValIleHisArgSer 367

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-471-780C-86

seq_documentation_block:
; Sequence 86, Application US/08471780C
; Patent No. 5759808
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,780C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,944
; FILING DATE: 17-AUG-1993
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,780C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,944
; FILING DATE: 17-AUG-1993
; APPLICATION NUMBER: FR 92402326.0
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400

```

```

; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-471-780C-86

alignment_scores:
    Quality: 31.00      Length: 6
    Ratio: 5.167       Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 83.333

alignment_block:
US-09-730-559B-45/rev x US-08-471-780C-86 ..

Align seg 1/1 to: US-08-471-780C-86 from: 1 to: 32

18 ATGAATAATATTAAAGCCC 1
17 MetAsnAsnLeuLysPro 22

seq_name: /cgn2_5/ptodata/1/iaa/5A_COMB.pep:US-08-467-282B-86

seq_documentation_block:
; Sequence 86, Application US/08467282B
; Patent No. 5800988
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,282B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,944
; FILING DATE: 17-AUG-1993
; APPLICATION NUMBER: FR 92402326.0
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-282B-86

```

alignment\_scores:  
 Quality: 31.00 Length: 6  
 Ratio: 5.167 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:  
 US-09-730-559B-45/rev x US-08-467-282B-86 ..

Align seg 1/1 to: US-08-467-282B-86 from: 1 to: 32

```

18 ATGAATAATATTAAAGCCC 1
|||||:|||||
17 MetAsnAsnLeuLysPro 22
  
```

seq\_name: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap:US-08-471-282A-86

```

seq_documentation_block:
; Sequence 86, Application US/08471282A
; Patent No. 5840853
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471.282A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106.944
; FILING DATE: 17-AUG-1993
; APPLICATION NUMBER: FR 92402326.0
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4400
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 86:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-471-282A-86
  
```

alignment\_scores:  
 Quality: 31.00 Length: 6  
 Ratio: 5.167 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:

```

US-09-730-559B-45/rev x US-08-471-282A-86 ..
Align seg 1/1 to: US-08-471-282A-86 from: 1 to: 32
18 ATGAATAATATTAAAGCCC 1
|||||:|||||
17 MetAsnAsnLeuLysPro 22

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pap:US-08-466-710C-86

seq_documentation_block:
; Sequence 86, Application US/08466710C
; Patent No. 5874541
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,710C
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/106,944
; FILING DATE: 17-AUG-1993
; APPLICATION NUMBER: FR 92402326.0
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4400
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-710C-86
  
```

alignment\_scores:  
 Quality: 31.00 Length: 6  
 Ratio: 5.167 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:

US-09-730-559B-45/rev x US-08-466-710C-86 ..

Align seg 1/1 to: US-08-466-710C-86 from: 1 to: 32

```

18 ATGAATAATATTAAAGCCC 1
|||||:|||||
17 MetAsnAsnLeuLysPro 22
  
```

```
seq_name: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:US-08-468-739C-86
seq_documentation_block:
; Sequence 86, Application US/08468739C
; Patent No. 6015695
; GENERAL INFORMATION:
; APPLICANT: Casterman, Cecile
; APPLICANT: Hamers, Raymond
; TITLE OF INVENTION: Immunoglobulins Devoid of Light Chains
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,739C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,944
; FILING DATE: 17-AUG-1993
; APPLICATION NUMBER: FR 92402326.0
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93401310.3
; FILING DATE: 21-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 04958.0008-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-739C-86

alignment_scores:
    Quality: 31.00      Length: 6
    Ratio: 5.167        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 83.333

alignment_block:
US-09-730-559B-45/rev x US-08-468-739C-86      ..
Align seg 1/1 to: US-08-468-739C-86 from: 1 to: 32

18 ATGAATAATATTAAAGCCC 1
|||||.....|
17 MetAsnAsnLeuLysPro 22

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-08-368-776A-3
seq_documentation_block:
; Sequence 3, Application US/08368776A
; Patent No. 6300482
; GENERAL INFORMATION:
; APPLICANT: Ciossek, Thomas
; APPLICANT: Ullrich, Axel
```

```
; APPLICANT: Millauer, Birgit
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS
; AND TREATMENT OF MDK1
; TITLE OF INVENTION: SIGNAL TRANSDUCTION
; DISORDERS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/368,776A
; FILING DATE: January 3, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-368-776A-3

alignment_scores:
    Quality: 31.00      Length: 7
    Ratio: 4.429        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 71.429

alignment_block:
US-09-730-559B-45/rev x US-08-368-776A-3      ..
Align seg 1/1 to: US-08-368-776A-3 from: 1 to: 610

21 TCTATGAATAATATTAAAGCCC 1
|||||.....|
501 SerIleAsnAsnLeuLysPro 507

seq_name: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:PCT-US96-00419-3
seq_documentation_block:
; Sequence 3, Application PC/TUS9600419
; GENERAL INFORMATION:
; APPLICANT: Thomas Ciossek, Axel Ullrich, Birgit
; APPLICANT: Millauer
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF
; MDK1 SIGNAL TRANSDUCTION DISORDERS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
```

```
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/368,776A
; FILING DATE: January 3, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: none
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 626 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-368-776A-5

alignment_scores:
  Quality: 31.00      Length: 7
  Ratio: 4.429      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 71.429

alignment_block:
  US-09-730-559B-45/rev x US-08-368-776A-5 ..
  Align seg 1/1 to: US-08-368-776A-5 from: 1 to: 626

      21 TCTATGAATAATATTAGCCC 1
         |||:::|||:::|||::||
      501 SerIleAsnLeuLysPro 507

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-368-776A-5

seq_documentation_block:
; Sequence 5, Application US/08368776A
; Patent No. 630482
; GENERAL INFORMATION:
; APPLICANT: Ciossek, Thomas
; APPLICANT: Ullrich, Axel
; APPLICANT: Millauer, Birgit
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF MDK1
; TITLE OF INVENTION: SIGNAL TRANSDUCTION
; TITLE OF INVENTION: DISORDERS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066

; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00419
; FILING DATE: January 3, 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: none
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 208/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US96-00419-3

alignment_scores:
  Quality: 31.00      Length: 7
  Ratio: 4.429      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 71.429

alignment_block:
  US-09-730-559B-45/rev x PCT-US96-00419-3 ..
  Align seg 1/1 to: PCT-US96-00419-3 from: 1 to: 610

      21 TCTATGAATAATATTAGCCC 1
         |||:::|||:::|||::||
      501 SerIleAsnLeuLysPro 507
```

---



```
24 CGATCTATGTAATAATATTAAAGCCC 1
   ::::::::::::::::::::
222 LysSerMetAsnValLysPro 229
seq_name: pir2:T32093

seq_documentation_block:
hypothetical protein K09F6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
C:Accession: T32093
R:Scheet, P.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid K09F6.
A:Reference number: Z21122
A:Accession: T32093
A:Molecule type: DNA
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-351 <SCH>
A:Cross-references: EMBL:AF016683; PIDN:AAB66192.1; GSPDB:GN00020; CESP:K09F6.1
A:Experimental source: strain Bristol N2; clone K09F6
C:Genetics:
A:Gene: CESP:K09F6.1
A:Map position: 2
C:Superfamily: Caenorhabditis elegans hypothetical protein K09F6.1

alignment_scores:
Quality: 38.00 Length: 8
Ratio: 4.750 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 75.000

alignment_block:
US-09-730-559B-45/rev x T32093 ..
Align seg 1/1 to: T32093 from: 1 to: 351
24 CGATCTATGTAATAATATTAAAGCCC 1
   ::::::::::::::::::::
263 LysSerMetAsnValLysPro 270
seq_name: pir1:A40811

seq_documentation_block:
myosin-light-chain kinase (EC 2.7.1.117) A - slime mold (Dictyostelium discoideum)
N:Alternate names: MLCK-A
C:Species: Dictyostelium discoideum
C:Date: 10-Apr-1992 #sequence_revision 21-Jan-1997 #text_change 11-Jun-1999
C:Accession: A40811; A37125
R:Tan, J.L.; Spudich, J.A.
J. Biol. Chem. 266, 16044-16049, 1991
A:Title: Characterization and bacterial expression of the Dictyostelium myosin light chain
A:Reference number: A40811; MUID:91340753
A:Accession: A40811
A:Molecule type: mRNA
A:Residues: 1-301 <TAN>
A:Cross-references: GB:M64176; NID:g1498249; PIDN:AAB06337.1; PID:g1498250
R:Tan, J.L.; Spudich, J.A.
J. Biol. Chem. 265, 13818-13824, 1990
A:Title: Dictyostelium myosin light chain kinase. Purification and characterization.
A:Reference number: A37125; MUID:90337997
A:Accession: A37125
A:Molecule type: protein
A:Residues: 9-12, 1', 14-19; 163-167, 'S', 169-179; 192-198 <TA2>
C:Genetics:
A:Gene: mlka
C:Function:
A:Description: catalyzes the formation of peptidyl-serine-phosphate or peptidyl-threonin
d in itself
A:Pathway: cytokinesis; fruiting body formation
A:Note: not activated by Ca2+/calmodulin in contrast with MLCK from higher eukaryotes
C:Superfamily: kinase-related transforming protein; protein kinase homology

C:Keywords: ATP; autophosphorylation; cell division; phosphoprotein; phosphotransfera
F:6-265/Domain: protein kinase homology <KIN>
F:14-22/Region: protein kinase ATP-binding motif
F:267-293/Domain: inhibitory #status predicted <INH>
F:37/Active site: Lys #status predicted
F:296/Binding site: phosphate (Thr) (covalent) (by autophosphorylation) #status predi

alignment_scores:
Quality: 36.00 Length: 7
Ratio: 5.143 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x A40811 ..
Align seg 1/1 to: A40811 from: 1 to: 301
1 GGGCTTAATATTATTCATAGA 21
   ::::::::::::::::::::
123 GlyLeuAsnIleValHisArg 129
seq_name: pir2:T21558

seq_documentation_block:
hypothetical protein F29G6.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T21558
R:Harris, B.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19441
A:Accession: T21558
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1785 <WIL>
A:Cross-references: EMBL:Z78543; PIDN:CAB01754.1; GSPDB:GN00028; CESP:F29G6.3b
A:Experimental source: clone F29G6
C:Genetics:
A:Gene: CESP:F29G6.3b
A:Map position: X
A:Introns: 10/1; 43/3; 61/3; 88/3; 110/3; 144/2; 183/3; 218/3; 252/3; 288/3; 1083/3

alignment_scores:
Quality: 35.00 Length: 8
Ratio: 4.375 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-730-559B-45 x T21558 ..
Align seg 1/1 to: T21558 from: 1 to: 1785
1 GGGCTTAATATTATTCATAGATCG 24
   ::::::::::::::::::::
261 GlyValAsnValLeuHisArgSer 268
seq_name: pir2:T21559

seq_documentation_block:
hypothetical protein F29G6.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T21559
R:Harris, B.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19441
A:Accession: T21559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1929 <WIL>
A:Cross-references: EMBL:Z78543; PIDN:CAB01755.1; GSPDB:GN00028; CESP:F29G6.3a
```



A:Experimental source: clone F29G6  
C:Genetics:  
A:Gene: CESP:F29G6.3a  
A:Map position: X  
A:Introns: 10/1; 43/3; 61/3; 88/3; 110/3; 144/2; 183/3; 218/3; 252/3; 288/3; 1083/3; 121

alignment\_scores:  
Quality: 35.00 Length: 8  
Ratio: 4.375 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-09-730-559B-45 x T21559 ..

Align seg 1/1 to: T21559 from: 1 to: 1929

1 GGGCTTAATATTATTCATAGATCG 24  
|||||:|||||:|||||:|||||  
261 GlyValAsnValLeuHisArgSer 268

seq\_name: pir2:B70335

seq\_documentation\_block:

conserved hypothetical protein aq\_389 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C:Accession: B70335

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

V. Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: B70335

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-182 <AQF>

A:Cross-references: GB:AE000688; NID:g2983063; PIDN:AAC06687.1; PID:g2983080; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: aq\_389

alignment\_scores:

Quality: 34.00 Length: 7  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x B70335 ..

Align seg 1/1 to: B70335 from: 1 to: 182

1 GGGCTTAATATTATTCATAGA 21  
|||||:|||||:|||||:|||||  
43 GlyLeuAsnIleIleHisLys 49

seq\_name: pir2:H90051

seq\_documentation\_block:

hypothetical protein [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: H90051

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: H90051

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <KUR>  
A:Cross-references: GB:BA000018; PID:g13702437; PIDN:BA843578.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA2275  
C:Superfamily: Staphylococcus aureus hypothetical protein lpl1

alignment\_scores:

Quality: 34.00 Length: 8  
Ratio: 4.250 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-09-730-559B-45 x H90051 ..

Align seg 1/1 to: H90051 from: 1 to: 258

2 GGCTTAATATTATTCATAGCGA 25  
|||||:|||||:|||||:|||||  
94 GlyLeuValLeuTyrValAspArg 101

seq\_name: pir2:F64238

seq\_documentation\_block:

hypothetical protein MG349 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 07-Dec-1999

C:Accession: F64238

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.;

M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,

, C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MUID:96026346

A:Accession: F64238

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-404 <TIGR>

A:Cross-references: GB:U39719; GB:L43967; NID:g1046055; PID:g1046056; TIGR:MG349

A:Experimental source: strain G-37

C:Genetics:

A:Genetic code: SGC3

alignment\_scores:

Quality: 34.00 Length: 7  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45/rev x F64238 ..

Align seg 1/1 to: F64238 from: 1 to: 404

21 TCTATGAATATATTAAGCC 1  
|||||:|||||:|||||:|||||  
383 SerLeuAsnIleLysPro 389

seq\_name: pir2:B87350

seq\_documentation\_block:

hypothetical protein CC0813 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: B87350

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete genome sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87350

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-540 <STO>  
A:Cross-references: EMBL:AE005673; NID:g13422062; PIDN:AAK22798.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0813

alignment\_scores:  
Quality: 34.00 Length: 8  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 87.500 Percent Identity: 87.500

alignment\_block:  
US-09-730-559B-45 x B87350 ..  
Align seg 1/1 to: B87350 from: 1 to: 540  
1 GGCTTAATATTCATAGATCGA 25  
|||||:|||||:|||||  
431 GlyLeuIleuPheTyAspArg 438

seq\_name: pir2:A11088

seq\_documentation\_block:  
transcription regulators, Fnr/Crp family homolog lmo0112 [imported] - Listeria monocytogenes  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: A11088  
R:Glaser, P.; Franguel, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.; Science 234, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: A11088  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-231 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC98327.1; PID:g16409471; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0112

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.714 Gaps: 0  
Percent Similarity: 87.500 Percent Identity: 75.000

alignment\_block:  
US-09-730-559B-45 x A11088 ..  
Align seg 1/1 to: A11088 from: 1 to: 231  
2 GGCTTAATATTCATAGATCGA 25  
||||:|||||:|||||  
96 GlyThrValLeuPheIleAspArg 103

seq\_name: pir2:T37924

seq\_documentation\_block:  
rna polymerase I specific transcription initiation on factor - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T37924  
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V. submitted to the EMBL Data Library, December 1995  
A:Reference number: Z21754  
A:Accession: T37924  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-599 <CON>  
A:Cross-references: EMBL:Z68198; PIDN:CAA92389.1; GSPDB:GN00066; SPDB:SPAC1866.11c  
A:Experimental source: strain 972h-; cosmid c1866  
C:Genetics:  
A:Gene: SPDB:SPAC1866.11c  
A:Map position: 1

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:  
US-09-730-559B-45 x T37924 ..  
Align seg 1/1 to: T37924 from: 1 to: 599  
1 GGCTTAATATTCATAGATCG 24  
|||||:|||||:|||||  
488 GlyLeuIleuLeuHisArgSer 495

seq\_name: pir2:B30239

seq\_documentation\_block:  
hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) II - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 05-May-2000  
C:Accession: B30239; B24317; S55972  
R:Basson, M.E.; Thorsness, M.; Finer-Moore, J.; Stroud, R.M.; Rine, J. Mol. Cell. Biol. 8, 3797-3808, 1988  
A:Title: Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl-CoA reductase  
A:Reference number: A93105; MUID:89127221  
A:Accession: B30239  
A:Molecule type: DNA  
A:Residues: 1-1045 <BAS>  
A:Cross-references: EMBL:M22255; NID:g171687; PIDN:AAA34677.1; PID:g171688  
R:Basson, M.E.; Thorsness, M.; Rine, J. Proc. Natl. Acad. Sci. U.S.A. 83, 5563-5567, 1986  
A:Title: Saccharomyces cerevisiae contains two functional genes encoding 3-hydroxy-3-methylglutaryl-CoA reductase  
A:Reference number: A94112; MUID:86287298  
A:Accession: B24317  
A:Molecule type: DNA  
A:Residues: 772-961 <BAS2>  
R:Dur, Z. submitted to the EMBL Data Library, March 1995  
A:Description: The sequence of S. cerevisiae cosmid 9324.  
A:Reference number: S55966  
A:Accession: S55972  
A:Molecule type: DNA  
A:Residues: 1-1045 <DUZ>  
A:Cross-references: GB:U22382; NID:g717059; PIDN:AAB67527.1; PID:g717061; MIPS:YLR450  
C:Genetics:  
A:Gene: SGD:HMG2  
A:Cross-references: SGD:S0004442; MIPS:YLR450w  
A:Map position: 12R  
C:Superfamily: hydroxymethylglutaryl-CoA reductase I  
C:Keywords: coenzyme A; oxidoreductase; sterol biosynthesis; transmembrane protein  
F:29-45/Domain: transmembrane #status predicted <TM1>  
F:248-264/Domain: transmembrane #status predicted <TM2>  
F:331-347/Domain: transmembrane #status predicted <TM3>  
F:402-418/Domain: transmembrane #status predicted <TM4>  
F:502-518/Domain: transmembrane #status predicted <TM5>  
F:681-697/Domain: transmembrane #status predicted <TM6>  
F:991-1007/Domain: transmembrane #status predicted <TM7>

alignment\_scores:  
Quality: 33.00 Length: 7  
Ratio: 4.714 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

```

alignment_block:
US-09-730-559B-45 x B30239  ..
Align seg 1/1 to: B30239 from: 1 to: 1045
      4 CTTAATATTATTCATAGATCG 24
      :|||||:|||||:|||||:
360 ileAsnIleIleHisArgSer 366

seq_name: pir2:G82595

seq_documentation_block:
ABC transporter phosphate binding protein XF2141 [imported] - Xylella fastidiosa (strain
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: G82595
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82595
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <SIM>
A:Cross-references: GB:AE004028; GB:AE003849; NID:g9107266; PIDN:AAF84940.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigri
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
C:Gene: XF2141
C:Superfamily: phosphate-repressible phosphate-binding protein

alignment_scores:
Quality: 32.00 Length: 7
Ratio: 4.571 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-730-559B-45 x G82595  ..
Align seg 1/1 to: G82595 from: 1 to: 364
      4 CTTAATATTATTCATAGATCG 24
      :|||||:|||||:|||||:
179 ileAsnIleValHisArgSer 185

seq_name: pir2:C87134

seq_documentation_block:
conserved hypothetical protein ML1802 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87134
R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

```

```

A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: C87134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <STO>
A:Cross-references: GB:AL450380; NID:g13093513; PIDN:CAC30755.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1802

```

```

alignment_scores:
Quality: 32.00 Length: 7
Ratio: 4.571 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x C87134  ..
Align seg 1/1 to: C87134 from: 1 to: 374
      5 TTAAATATTATTCATAGATCGA 25
      :|||||:|||||:|||||:
54 LeuValIeuPheIleAspArg 60

```

---

OM of: US-09-730-559b-45 to: SwissProt\_40:\* out\_format : pfs  
 Date: Sep 8, 2002 8:13 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL=frame.n2p.model -DEV=xlh  
 -Q/cgn2\_1/USPTO\_spoal/US09730559/runat\_03092002\_150446\_3360/app\_query.fasta\_1.86  
 -DB=SwissProt\_40 -QPMT=fastan -SUFFIX=rsd -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
 -USR=US09730559\_@CNCI\_1\_13 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-730-559b-45  
 Query length: 26  
 Database: SwissProt\_40:\*  
 Database sequences: 105224  
 Database length: 38719550  
 Search time (sec): 28.210000

score_list:	Sequence	Strd Orig	zScore	EScore	Len	! Documentation
	SwissProt_40:YMD4_CAEEL -	38.00	136.67	1.89	343	P34458 caenorhabditis elegans
	SwissProt_40:KMLC_DICDI +	36.00	129.84	5.26	295	P25323 dictyostelium discoideum
	SwissProt_40:Y349_MYCGE -	34.00	119.40	14.66	404	P47591 mycoplasma genitalium
	SwissProt_40:YAGB_SCHPO +	33.00	112.34	24.46	599	P01010 schizosaccharomyces pombe
	SwissProt_40:HMD2_YEAST +	33.00	108.01	24.44	1045	P12684 saccharomyces cerevisiae
	SwissProt_40:OXDA_MOUSE +	32.00	112.61	40.89	346	P18894 mus musculus (mouse)
	SwissProt_40:OXDA_MOUSE +	32.00	112.22	40.88	364	P09633 xylella fastidiosa
	SwissProt_40:PSTS_XYFLA +	32.00	110.76	40.87	439	P03330 saccharomyces cerevisiae
	SwissProt_40:GCN5_YEAST +	32.00	103.94	40.80	1054	P12683 saccharomyces cerevisiae
	SwissProt_40:HMD1_YEAST +	32.00	103.94	40.80	1054	P12683 saccharomyces cerevisiae
	SwissProt_40:YHVS_YEAST +	32.00	102.75	40.79	1228	P13851 saccharomyces cerevisiae
	SwissProt_40:RS7_MICLU -	31.00	114.81	68.37	156	P09898 microcococcus luteus (m
	SwissProt_40:MORE_BRUME +	31.00	114.47	68.37	163	P08966 brucella melitensis
	SwissProt_40:RL15_QUESE +	31.00	112.84	68.34	201	P08212 quercus suber (cork o
	SwissProt_40:Y008_METJA -	31.00	112.14	68.33	220	P060319 methanococcus jannasch
	SwissProt_40:Y007_BPP22 -	31.00	111.83	68.32	229	P01074 bacteriophage p22. dn
	SwissProt_40:FAEL_ECOLI +	31.00	110.93	68.31	257	P13784 escherichia coli. k88
	SwissProt_40:YQ23_CAEEL +	31.00	108.20	68.27	365	P09452 caenorhabditis elegans
	SwissProt_40:Y40G_RHISN +	31.00	106.60	68.24	448	P55628 rhizobium sp. (strain
	SwissProt_40:IMDH_PYRAB +	31.00	105.99	68.23	485	P09449 pyrococcus abyssi. in
	SwissProt_40:IMDH_PYRAB +	31.00	105.99	68.23	485	P42851 pyrococcus furiosus
	SwissProt_40:IMDH_PYRHO +	31.00	105.97	68.23	486	P08045 pyrococcus horikoshii
	SwissProt_40:IMDH_PYRHO +	31.00	105.81	68.23	496	P059011 methanococcus jannasch
	SwissProt_40:AMPA_RICPR +	31.00	105.75	68.23	500	P27888 rickettsia prowazekii
	SwissProt_40:GUAA_ECOLI +	31.00	105.37	68.22	525	P04079 escherichia coli. gnd
	SwissProt_40:YMXI_CAEEL +	31.00	105.05	68.22	547	P34509 caenorhabditis elegans
	SwissProt_40:FRPL_SCHPO +	31.00	104.81	68.21	564	P04800 schizosaccharomyces p
	SwissProt_40:NODO_AZOBR +	31.00	104.07	68.20	620	P28604 a nodq bifunctional e
	SwissProt_40:SMX2_YEAST +	31.00	102.19	68.17	790	P132909 saccharomyces cerevis
	SwissProt_40:CLP_TRYBB +	31.00	101.46	68.16	868	P13543 trypanosoma brucei br
	SwissProt_40:CLPB_TREPA +	31.00	101.37	68.16	878	P083110 treponema pallidum. d
	SwissProt_40:EPAT_CHICK -	31.00	100.41	68.14	993	P04242 gallus gallus (chicken
	SwissProt_40:EPAT_HUMAN -	31.00	100.37	68.14	998	P15375 homo sapiens (human)
	SwissProt_40:EPAT_MOUSE -	31.00	100.37	68.14	998	P061772 mus musculus (mouse)
	SwissProt_40:EPAT_RAT -	31.00	100.37	68.14	998	P061772 mus musculus (mouse)
	SwissProt_40:CALF_HUMAN +	31.00	96.68	68.08	1603	P547599 rattus norvegicus (rat)
	SwissProt_40:YLU9_CAEEL -	30.00	112.29	114.21	129	P107092 homo sapiens (human)
	SwissProt_40:YLU9_CAEEL -	30.00	112.29	114.21	129	P134403 caenorhabditis elegans
	SwissProt_40:RS7_MYCSE -	30.00	110.87	114.17	155	P30764 mycobacterium leprae.
	SwissProt_40:RS7_MYCSE -	30.00	110.87	114.17	155	P41193 mycobacterium smegmati
	SwissProt_40:RS7_MYCTO -	30.00	110.87	114.17	155	P41194 mycobacterium tubercul
	SwissProt_40:RS7_MYCBO -	30.00	110.82	114.17	156	P53539 mycobacterium bovis. 3

SwissProt_40:RS7_STRCO -	30.00	110.82	114.17	156	P0910k4 streptomyces coelic
SwissProt_40:MOAE_RHILO +	30.00	110.67	114.17	159	P0984p0 rhizobium loti (me
SwissProt_40:VP11_RDVA +	30.00	109.66	114.14	181	P28890 rice dwarf virus (i
SwissProt_40:VP11_RDVF +	30.00	109.32	114.13	189	P085442 rice dwarf virus (i
SwissProt_40:Y278_ARATH +	30.00	107.63	114.09	235	P082246 arabidopsis thaliana

seq\_name: SwissProt\_40:YMD4\_CAEEL

seq\_documentation\_block:  
 ID YMD4\_CAEEL STANDARD; PRT; 343 AA.  
 AC P34458;

DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 40.8 kDa protein F54H12.4 in chromosome III.  
 GN F54H12.4.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,  
 Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
 Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
 Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K.,  
 Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 Wierston P.;  
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).

EMBL; L25599; AAA28053.1; -

PIR; S44834; S44834.

DR WormPep: F54H12.4; CE00546.

KW Hypothetical protein.

SQ SEQUENCE 343 AA; 40799 MW; F636DCB35BA96F96 CRC64;

alignment\_scores:  
 Quality: 38.00 Length: 8  
 Ratio: 4.750 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45/rev x YMD4\_CAEEL  
 Align seg 1/1 to: YMD4\_CAEEL from: 1 to: 343

24 CGATCTATGAATATATTAAGCCC 1

222 LysSerMetAsnValLysPro 229

seq\_name: SwissProt\_40:KMLC\_DICDI

seq\_documentation\_block:

ID KMLC\_DICDI STANDARD; PRT; 295 AA.  
 AC P25323;

DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE MYOSIN light chain kinase (EC 2.7.1.117) (MLCK).  
GN MLCK.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX3;  
RX MEDLINE=91340753; PubMed=1651931;  
RA Tan J.L., Spudich J.A.;  
RT "Characterization and bacterial expression of the Dictyostelium  
myosin light chain kinase cDNA. Identification of an autoinhibitory  
domain.";  
RL J. Biol. Chem. 266:16044-16049(1991).  
[2]  
RP REVISIONS.  
RA Spudich J.A.;  
RN Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
[3]  
RN PARTIAL SEQUENCE.  
RC STRAIN=AX3;  
RX MEDLINE=90337997; PubMed=2380188;  
RA Tan J.L., Spudich J.A.;  
RT "Dictyostelium myosin light chain kinase. Purification and  
characterization.";  
RL J. Biol. Chem. 265:13818-13824(1990).  
CC -1- FUNCTION: PHOSPHORYLATES A SPECIFIC SERINE IN THE N-TERMINUS OF A  
MYOSIN LIGHT CHAIN.  
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin  
light-chain] phosphate.  
CC -1- ENZYME REGULATION: POSSESSES AN AUTOINHIBITORY DOMAIN.  
CC AUTOPHOSPHORYLATION APPEARS TO INCREASE THE ENZYMIC ACTIVITY.  
CC DOES NOT HAVE A CALMODULIN-BINDING DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC CAMK SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M64176; AAB06337.1; -  
DR PIR; A40811; A40811.  
DR PIR; A37125; A37125.  
DR HSP; Q63450; 1A06.  
DR DictyDb; DD01034; mlka.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR SMART; SM00220; S\_TRC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Phosphorylation.  
FT DOMAIN 8 265 PROTEIN KINASE.  
FT DOMAIN 264 295 AUTOINHIBITORY DOMAIN.  
FT NP\_BIND 14 22 ATP (BY SIMILARITY).  
FT BINDING 37 37 ATP (BY SIMILARITY).  
FT ACT\_SITE 130 130 BY SIMILARITY.  
SQ SEQUENCE 295 AA; 33406 MW; 546CAED8F6ECD0B CRC64;  
  
alignment\_scores:  
Quality: 36.00 Length: 7  
Ratio: 5.143 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:  
US-09-730-559B-45 x KMLC\_DICDI ..  
  
Align seg 1/1 to: KMLC\_DICDI from: 1 to: 295  
  
1 GGGCTTAATATATTATTCATAGA 21  
|||||  
123 GlyLeuAsnIleValHisArg 129  
  
seq\_name: SwissProt\_40:Y349\_MYCGE  
  
seq\_documentation\_block:  
ID Y349\_MYCGE STANDARD; PRT; 404 AA.  
AC P47591;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MG349.  
GN MG349.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569693;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
Nguyen D.T., Weidback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U39716; AAC71574.1; -  
DR TIGR; MG349; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 404 AA; 47774 MW; 37023D6A8F92CAE4 CRC64;  
  
alignment\_scores:  
Quality: 34.00 Length: 7  
Ratio: 4.857 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714  
  
alignment\_block:  
US-09-730-559B-45/rev x Y349\_MYCGE ..  
  
Align seg 1/1 to: Y349\_MYCGE from: 1 to: 404  
  
21 TCTATGAATAATTAAGCC 1  
|||||  
383 SerLeuAsnIleValHisPro 389  
  
seq\_name: SwissProt\_40:YAOB\_SCHPO  
  
seq\_documentation\_block:  
ID YAOB\_SCHPO STANDARD; PRT; 599 AA.  
AC Q10110;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 68.6 kDa protein C18G6.11c in chromosome I.  
GN SPAC18G6.11c.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO YEAST RRN3 AND C.ELEGANS C36E8.1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z68198; CAA92389.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 599 AA; 68554 MW; 8351E6612CE768EF CRC64;  
  
alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000  
  
alignment\_block:  
US-09-730-559B-45 x YAOB\_SCHPO ..  
  
Align seq 1/1 to YAOB\_SCHPO from: 1 to: 599  
  
1 GGGCTTAATATTCATGATCG 24  
|||||:||||:|||||  
488 GlyLeuGlulleuHisArgSer 495  
  
seq\_name: SwissProt\_40:HMD2\_YEAST  
  
seq\_documentation\_block:  
ID\_HMD2\_YEAST STANDARD; PRT; 1045 AA.  
AC P12684;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase 2 (EC 1.1.1.34) (HMG-  
DE CoA reductase 2).  
GN HMG2 OR YLR450W OR L9324.2.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89127221; PubMed=3065625;  
RA Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J.;  
RT "Structural and functional conservation between yeast and human 3-  
RT hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting  
RT enzyme of sterol biosynthesis."  
RL Mol. Cell. Biol. 8:3797-3808(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J.,  
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
RA Johnson D., Johnston Y., Langston J., Latreille P., Le T.,  
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
RA Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,  
  
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 772-961 FROM N.A.  
RX MEDLINE=86287298; PubMed=3526336;  
RA Basson M.E., Thorsness M., Rine J.;  
RT "Saccharomyces cerevisiae contains two functional genes encoding 3-  
RT hydroxy-3-methylglutaryl-coenzyme A reductase."  
RL Proc. Natl. Acad. Sci. U.S.A. 83:5563-5567(1986).  
CC -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE  
CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING  
CC ENZYME OF THE STEROL BIOSYNTHESIS.  
CC -1- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) = (S)-3-  
CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.  
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
CC reticulum.  
CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; M22255; AAA34677.1; -;  
DR EMBL; U22382; AAB67527.1; -;  
DR PIR; B24317; B24317.  
DR PIR; B30239; B30239.  
DR SGD; S000442; HMG2.  
DR InterPro; IPR002202; HMG-CoA\_red; 1.  
DR Pfam; PF00368; HMG-CoA\_red; 1.  
DR PRINTS; PR00071; HMGCOARDTASE.  
DR PROSITE; PS00066; HMG\_COA\_REDUCTASE\_1; 1.  
DR PROSITE; PS00318; HMG\_COA\_REDUCTASE\_2; 1.  
DR PROSITE; PS01192; HMG\_COA\_REDUCTASE\_3; 1.  
DR PROSITE; PS50065; HMG\_COA\_REDUCTASE\_4; 1.  
DR PROSITE; PS50156; SSD; 1.  
KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;  
KW Cholesterol biosynthesis; NADP; Multigene family.  
FT DOMAIN 1 523 MEMBRANE-BOUND.  
FT DOMAIN 524 613 LINKER.  
FT DOMAIN 614 1045 CATALYTIC.  
FT TRANSMEM 27 53 POTENTIAL.  
FT TRANSMEM 186 210 POTENTIAL.  
FT TRANSMEM 241 265 POTENTIAL.  
FT TRANSMEM 299 323 POTENTIAL.  
FT TRANSMEM 331 356 POTENTIAL.  
FT TRANSMEM 397 421 POTENTIAL.  
FT TRANSMEM 498 523 POTENTIAL.  
FT ACT\_SITE 710 710 BY SIMILARITY.  
FT ACT\_SITE 920 920 BY SIMILARITY.  
FT ACT\_SITE 1016 1016 GENERAL BASE (BY SIMILARITY).  
FT CARBOHYD 115 115 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 150 150 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 158 158 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 179 179 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 428 428 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1045 AA; 115691 MW; 1FD9DCD3AC01B15E CRC64;  
  
alignment\_scores:  
Quality: 33.00 Length: 7  
Ratio: 4.714 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x HMD2\_YEAST

Align seg 1/1 to: HMD2\_YEAST from: 1 to: 1045

4 CTTAATATTATTCATAGACG 24

:::|||||

360 IleAsnLeuIleHisArgSer 366

seq\_name: SwissProt\_40:OXDA\_MOUSE

## seq\_documentation\_block:

ID OXDA\_MOUSE STANDARD; PRT; 346 AA.

AC P18894; 064465;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE D-amino acid oxidase (EC 1.4.3.3) (DAMO) (DAAO).

GN DAO OR DAO1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=90382679; PubMed=1976103;

RA Tada M., Fukui K., Momoi K., Miyake Y.;

RT "Cloning and expression of a cDNA encoding mouse kidney D-amino acid

RT oxidase.";

RL Gene 90:293-297(1990).

RN [2]

RP SEQUENCE FROM N.A.

RA Sasaki M.;

RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.

RN [3]

RP MUTAGENESIS OF GLY-182.

RX MEDLINE=92385496; PubMed=1355365;

RA Sasaki M., Konno R., Nishio M., Niwa A., Yasumura Y., Enami J.;

RT "A single-base-pair substitution abolishes D-amino-acid oxidase

RT activity in the mouse.";

RL Biochim. Biophys. Acta 1139:315-318(1992).

CC -1- FUNCTION: COULD ACT AS A DETOXIFYING AGENT WHICH REMOVES D-AMINO

CC ACIDS ACCUMULATED DURING AGING. ACTS ON A VARIETY OF D-AMINO ACIDS

CC WITH A PREFERENCE FOR THOSE HAVING SMALL HYDROPHOBIC SIDE CHAINS

CC FOLLOWED BY THOSE BEARING POLAR, AROMATIC, AND BASIC GROUPS. DOES

CC NOT ACT ON ACIDIC AMINO ACIDS.

CC -1- CATALYTIC ACTIVITY: A D-amino acid + H(2)O + O(2) -&gt; a 2-oxo acid +

CC NH(3) + H(2)O(2).

CC -1- COFACTOR: FAD.

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Peroxisomal.

CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M32209; AAA39367.1; -

CC EMBL; D10210; BAA01062.1; -

CC EMBL; D10211; BAA01063.1; -

CC PIR; JH0185; JH0185.

CC HSSP; P00371; 1AN9.

CC MGD; MGI:94859; Dao1.

CC InterPro; IPR000927; DAO.

CC Pfam; PF01466; DAO; 1.

CC PROSITE; PS00342; MICROBODIES\_CTER; 1.

CC PROSITE; PS00677; DAO; 1.

CC Oxidoreductase; Flavoprotein; FAD; Peroxisome.

KW

FT NP\_BIND 3 17 FAD (ADP PART) (POTENTIAL).  
FT ACT\_SITE 227 227 BY SIMILARITY.  
FT ACT\_SITE 306 306 BY SIMILARITY.  
FT SITE 344 346 MICROBODY TARGETING SIGNAL.  
FT MUTAGEN 182 182 G->R: ABOLISHES ACTIVITY.  
FT CONFLICT 64 64 A -> V (IN REF. 1).  
FT CONFLICT 157 157 N -> K (IN REF. 1).  
FT CONFLICT 173 173 MISSING (IN REF. 1).  
SQ SEQUENCE 346 AA; 38714 MW; DFD18340675081C1 CRC64;

## alignment\_scores:

Quality: 32.00 Length: 7

Ratio: 4.571 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 71.429

## alignment\_block:

US-09-730-559B-45 x OXDA\_MOUSE

Align seg 1/1 to: OXDA\_MOUSE from: 1 to: 346

1 GGCCTTAATATTATTCATAGA 21

|||||

155 GlyValAsnLeuIleHisArg 161

seq\_name: SwissProt\_40:PSTS\_XYLFA

## seq\_documentation\_block:

ID PSTS\_XYLFA STANDARD; PRT; 364 AA.

AC Q9PBK3;

DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Phosphate-binding periplasmic protein precursor (PBP).

GN PSTS OR XP2141.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;

OC Xylella.

OX NCBI\_TaxID=2371;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=9a5c;

RX MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,

RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,

RA Coutinho L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,

RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,

RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitchima J.P.,

RA Krieger J.E., Kuranai E.E., Laigret F., Lambais M.R., Leite L.C.C.,

RA Lenos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

RA Menck C.F.M., Miraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,

RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,

RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,

RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

RA de Souza A.P., Terenzi M.A., Truffi D., Tsai S.M., Tshako M.H.,

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).

CC -1- FUNCTION: REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE

CC TRANSPORT (BY SIMILARITY).

CC



```

CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004028; AAF84940.1; -.
DR HSSP; P06128; 11XH.
DR InterPro; IPR002370; PSTs.
DR Pfam; PF01449; PSTs; 1.
KW Phosphate transport; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 364 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 364 AA; 38736 MW; 8609CFAA159D4277 CRC64;

alignment_scores:
  Quality: 32.00 Length: 7
  Ratio: 4.571 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-730-559B-45 x PSTS_XYLFA ..

Align seg 1/1 to: PSTS_XYLFA from: 1 to: 364

4 CTTAATATATTCATGATCG 24
:::|||||:::|||||
179 ILeAsnllvAlHIsArgSer 195

seq_name: SwissProt_40:GCN5_YEAST

seq_documentation_block:
ID GCN5_YEAST STANDARD; PRT; 439 AA.
AC Q03330;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Transcriptional activator GCN5.
GN GCN5 OR ADA4 OR YGR252W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93011009; PubMed-1396595;
RA Georgakopoulos T., Thireos G.;
RT "Two distinct yeast transcriptional activators require the function
RT of the GCN5 protein to promote normal levels of transcription.";
RL EMBO J. 11:4145-4152(1992).
RN [2]
RP SEQUENCE OF 1-170 FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE-97279234; PubMed-9133742;
RA Peroli F., Carignani G., Pavanello A., Guerreiro P., Azevedo D.,
RA Rodrigues-Pousada C., Melchiorretto P., Panzeri L.,
RA Agostoni Carbone M.L.;
RT "Analysis of a 17.9 kb region from Saccharomyces cerevisiae
RT chromosome VII reveals the presence of eight open reading frames,
RT including BRFL1 (TFIIB70) and GCN5 genes.";
RL Yeast 13:373-377(1997).
RN [3]
RP SEQUENCE OF 170-439 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE-97279233; PubMed-9133741;
RA Mazzoni C., Ruzzi M., Rinaldi T., Solinas F., Montebove F.,

```

```

RA Frontali L.;
RT "Sequence analysis of a 10.5 kb DNA fragment from the yeast
RT chromosome VII reveals the presence of three new open reading frames
RT and of a tRNAThr gene.";
RL Yeast 13:369-372(1997).
RN [4]
RP ASSOCIATION WITH ADA2.
RX MEDLINE-95045371; PubMed-7957049;
RA Marcus G.A., Silverman N., Berger S.L., Horiuchi J., Guarente L.;
RT "Functional similarity and physical association between GCN5 and
RT ADA2: putative transcriptional adaptors.";
RL EMBO J. 13:4807-4815(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 99-262.
RX MEDLINE-99362688; PubMed=10430873;
RA Trievel R.C., Rojas J.R., Sterner D.E., Venkataramani R.N., Wang L.,
RA Zhou J., Allis C.D., Berger S.L., Marmorstein R.;
RT "Crystal structure and mechanism of histone acetylation of the yeast
RT GCN5 transcriptional coactivator.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8931-8936(1999).
CC -!- FUNCTION: GENERAL TRANSCRIPTIONAL ACTIVATOR OPERATING IN CONCERT
CC WITH CERTAIN OTHER DNA-BINDING TRANSCRIPTIONAL ACTIVATORS SUCH AS
CC GCN4 OR HAP2/3/4. FUNCTION AS AN HISTONE ACETYLTRANSFERASE (HAT)
CC TO PROMOTE TRANSCRIPTIONAL ACTIVATION. HAS A STRONG PREFERENCE FOR
CC LYSINE 14 OF H3 AND A SOMEWHAT LOWER PREFERENCE FOR LYSINES 8 AND
CC 16 OF HISTONE H4.
CC -!- SUBUNIT: HETERODIMER WITH ADA2. PART OF THE ADA/GCN5 COMPLEX THAT
CC CONSISTS OF HFI1/ADAL, ADA2, ADA3, SPI20/ADA5 AND GCN5.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68628; CAA48602.1; -.
DR EMBL; X73037; CAA97281.1; -.
DR EMBL; X99228; CAA67614.1; -.
DR PIR; S28051; S28051.
DR PDB; 1VGH; 02-AUG-99.
DR TRANSFAC; T02145; -.
DR SGD; S0003484; GCN5.
DR InterPro; IPR000182; Acetyltransf_GCN5.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00583; Acetyltransf; 1.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00144; BROMODOMAIN_2; 1.
KW Transcription regulation; DNA-binding; Activator; Trans-acting factor;
KW 3D-structure.
FT ACT_SITE 173 173 GENERAL BASE.
FT DOMAIN 344 414 BROMODOMAIN.
SQ SEQUENCE 439 AA; 51069 MW; 3200730DDC7EF70D CRC64;

alignment_scores:
  Quality: 32.00 Length: 8
  Ratio: 4.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-730-559B-45/rev x GCN5_YEAST ..

Align seg 1/1 to: GCN5_YEAST from: 1 to: 439

24 CGATCTATGAATAATATTAAAGCCC 1
:::|||||:::|||||

```

```
295 LysAspLeuAsnAsnIleLysPro 302
seq_name: SwissProt_40:HMD1_YEAST
seq_documentation_block:
ID HMD1_YEAST STANDARD; PRT; 1054 AA.
AC P12683;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 (EC 1.1.1.34) (HMG-
DE CoA reductase 1).
GN HMG1 OR YML075C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69127221; PubMed=3065625;
RA Basson M.E., Thorsness M., Finer-Moore J., Stroud R.M., Rine J.;
RT "Structural and functional conservation between yeast and human 3-
RT hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting
RT enzyme of sterol biosynthesis.";
RL Mol. Cell. Biol. 8:3797-3808(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 776-965 FROM N.A.
RX MEDLINE=86287298; PubMed=3526336;
RA Basson M.E., Thorsness M., Rine J.;
RT "Saccharomyces cerevisiae contains two functional genes encoding 3-
RT hydroxy-3-methylglutaryl-coenzyme A reductase.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5563-5567(1986).
CC -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
CC ENZYME OF THE STEROL BIOSYNTHESIS.
CC -1- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) = (S)-3-
CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DE EMBL; M22002; AAA34676.1; -
DR EMBL; Z46373; CA866503.1; -
DR PIR; A24317; A24317.
DR PIR; A30239; A30239.
DR PIR; S48822; S48822.
DR SGD; S0004540; HMG1.
DR InterPro; IPR002202; HMG-CoA_red.
DR InterPro; IPR000731; HMGCR_patched_5TM.
DR Pfam; PF00368; HMG-CoA_red; 1.
DR PRINTS; PR00071; HMGCOARDTASE.
DR PROSITE; PS00066; HMG_COA_REDUCTASE_1; 1.
DR PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
DR PROSITE; PS01192; HMG_COA_REDUCTASE_3; 1.
DR PROSITE; PS00065; HMG_COA_REDUCTASE_4; 1.
DR PROSITE; PS01556; SSD; 1.
KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
KW Cholesterol biosynthesis; NADP; Multigene family.
FT DOMAIN 1 524 MEMBRANE-BOUND.

FT DOMAIN 525 617 LINKER.
FT DOMAIN 618 1054 CATALYTIC.
FT TRANSMEM 27 53 POTENTIAL.
FT TRANSMEM 187 211 POTENTIAL.
FT TRANSMEM 242 266 POTENTIAL.
FT TRANSMEM 300 324 POTENTIAL.
FT TRANSMEM 332 357 POTENTIAL.
FT TRANSMEM 398 422 POTENTIAL.
FT TRANSMEM 499 524 BY SIMILARITY.
FT ACT_SITE 714 714 BY SIMILARITY.
FT ACT_SITE 924 924 BY SIMILARITY.
FT ACT_SITE 1020 1020 GENERAL BASE (BY SIMILARITY).
FT CARBOHYD 115 115 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1054 AA; 115625 MW; 2B624944FB7B2DD0 CRC64;

alignment_scores:
Quality: 32.00 Length: 7
Ratio: 4.571 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-730-559b-45 x HMD1_YEAST ..
Align seg 1/1 to: HMD1_YEAST from: 1 to: 1054

4 CTTAATATTATTCATAGATCG 24
:::||||:|||||||
361 MetAsnValIleHisArgSer 367

seq_name: SwissProt_40:YHV5_YEAST
seq_documentation_block:
ID YHV5_YEAST STANDARD; PRT; 1228 AA.
AC P38851;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 143.6 kDa protein in SPO16-REC104 intergenic region.
GN YHR155W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: STRONG, TO YEAST SIP3. SOME, TO S.POMBE SPAC19A8.02.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

Quality: 31.00 Length: 8
Ratio: 3.875 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-730-559B-45/rev x RS7_MICLU ..

Align seg 1/1 to: RS7_MICLU from: 1 to: 156

24 CGATCTATGAATAAATATTAAAGCC 1
      :::::|||||:|||||
64 LysAlaMetaspasInileysPro 71

seq_name: SwissProt_40:MOAE_BRUME

seq_documentation_block:
ID MOAE_BRUME STANDARD; PRT; 163 AA.
AC Q8YGA6;
DT 01-MAR-2002 (Rel. 41, Created)
DE 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Molybdopter in converting factor subunit 2 (MPT synthase subunit 2)
DE Molybdopter in synthase subunit 2 (Molybdenum cofactor biosynthesis
DE protein E) (Molybdopter in converting factor large subunit).
DE MOAE OR BMEII254.
GN Brucella melitensis.
OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Ivanova N., Anderson I., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Delvecchio V.G., Kapatral V., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablon E., Larsen N., D'Souza M., Bernal A., Mazur M., Goldstein E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- FUNCTION: Converts molybdopter in precursor 2 into molybdopter in.
CC This requires the incorporation of two sulfur atoms into precursor
CC 2 to generate the dithiolene group (By similarity).
CC -!- PATHWAY: Molybdenum cofactor biosynthesis.
CC -!- SUBUNIT: Heterodimer of 2 moad subunits and 2 moaE subunits (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE MOAE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE009564; AAL52435.1;
KW Molybdenum cofactor biosynthesis; Complete proteome.
SQ SEQUENCE 163 AA; 18396 MW; B622143ED5116118 CRC64;

alignment_scores:
Quality: 31.00 Length: 7
Ratio: 4.429 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 85.714

alignment_block:
US-09-730-559B-45 x MOAE_BRUME ..

Align seg 1/1 to: MOAE_BRUME from: 1 to: 163

1 GGGCTTAATATTATTATCATAGA 21

```

```

|||||:|||||:|||||:|||||
82 GlyLeuThrIleHisArg 88
seq_name: SwissProt_40:RL15_QUESU
seq_documentation_block:
ID RL15_QUESU STANDARD; PRT; 201 AA.
AC O82712;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L15.
GN RPL15.
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fagales; Fagaceae; Quercus.
OX NCBI_TaxID=58331;
RN [1]
RP SEQUENCE FROM N.A.
RA Huguet G., Pla M., Verdaguier D., Molinas M.;
RT "Ribosomal proteins in Quercus suber.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ001346; CAA04690.1; -
DR InterPro: IPR000439; Ribosomal L15e.
DR Pfam: PF00827; Ribosomal L15e; 1.
DR PROSITE: PS01194; RIBOSOMAL_L15E; 1.
KW Ribosomal protein.
SQ SEQUENCE 201 AA; 23135 MW; F77F66B61886CF3D CRC64;

alignment_scores:
Quality: 31.00 Length: 7
Ratio: 4.429 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-730-559B-45 x RL15_QUESU ..
Align seg 1/1 to: RL15_QUESU from: 1 to: 201

4 CTTAATATTATTCATAGTCG 24
|||||:|||||:|||||:|||||
33 LeuAsnValIleHisArgAla 39
seq_name: SwissProt_40:Y008_METJA
seq_documentation_block:
ID Y008_METJA STANDARD; PRT; 220 AA.
AC Q60319;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0008.
GN MJ0008.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; PubMed=8688087;
RX

```

```

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen A.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: SOME, TO M.JANNASCHII MJ1311.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67460; AAB97997.1; -
DR TIGR; MJ0008; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 220 AA; 25836 MW; 955DF4A0FAB71B88 CRC64;

alignment_scores:
Quality: 31.00 Length: 8
Ratio: 3.875 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 50.000

alignment_block:
US-09-730-559B-45/rev x Y008_METJA ..
Align seg 1/1 to: Y008_METJA from: 1 to: 220

24 CGATCTATGTAATATTAAGCCC 1
|||||:|||||:|||||:|||||
175 LysAsnLeuAsnAsnLeuLysPro 182
seq_name: SwissProt_40:VG07_BPP22
seq_documentation_block:
ID VG07_BPP22 STANDARD; PRT; 229 AA.
AC Q01074;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA transfer protein gp7 precursor.
GN 7.
OS Bacteriophage P22.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
OX NCBI_TaxID=10754;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92394890; PubMed=1522065;
RA Conlin C.A., Vimr E.R., Miller C.G.;
RT "Oligopeptidase A is required for normal phase P22 development.";
RL J. Bacteriol. 174:5869-5880(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Kropinski A.M.B., VanderByl C.S.;
RT "The completed sequence of genome of Salmonella phage P22.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 225-229 FROM N.A.
RX MEDLINE=93219140; PubMed=8464750;
RA Adhikari P., Berget P.B.;
RT "Sequence of a DNA injection gene from Salmonella typhimurium phage
RT P22.";
RL Nucleic Acids Res. 21:1499-1499(1993).
RN [4]

```

RP SEQUENCE OF 21-31.  
RX MEDLINE-91306435; PubMed-1853558;  
RA Eppler K., Wyckoff E., Goates J., Parr R., Casjens S.;  
RT "Nucleotide sequence of the bacteriophage P22 genes required for DNA  
packaging";  
RL Virology 183:519-538(1991).  
CC -!- FUNCTION: REQUIRED FOR EJECTION OF THE PHAGE DNA FROM THE PHAGE  
CC PARTICLE AND INJECTION OF THAT DNA INTO THE HOST.  
CC -!- SIMILARITY: STRONG, TO PHAGE APSE-1 P32.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M93985; AAA72115.1; -.  
DR EMBL; AF217253; AAF75053.1; -.  
DR EMBL; L07556; CAB23820.1; -.  
DR PIR; C43330; C43330.  
KW Late protein.  
FT PROPEP 1 20 REMOVED IN MATURE FORM.  
FT CHAIN 21 229 DNA TRANSFER PROTEIN GP7.  
SQ SEQUENCE 229 AA; 23407 MW; 051A26642F185982 CRC64;

## alignment\_scores:

Quality:	31.00	Length:	6
Ratio:	5.167	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	83.333

## alignment\_block:

US-09-730-559B-45/rev x VG07\_BPP22 ..

Align seg 1/1 to: VG07\_BPP22 from: 1 to: 229

18	ATGAATAATATTAGCCC	1
53	MetAsnAsnLeuLysPro	58



OM of: US-09-730-559b-45 to: SPTRMBL\_19:\* out\_format : pfs

Date: Sep 8, 2002 8:12 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet n2p.model -DEV=xlh
-Q/cn2_1/USPTO_SPOOL/US09730559/runat_03092002_150446_3306/app_query.fasta_1.86
-DB=SPTRMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=ptc
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09730559 -CGN1_1.60 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT_THREADS=1
```

Search information block:

```
Query: US-09-730-559b-45
Query length: 26
Database: SPTRMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 84.610000
```

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
sp_invertebrate:018242	-	38.00	136.94	8.13	343	018242 caenorhabditis elegans
sp_invertebrate:016733	-	38.00	136.76	8.14	351	016733 caenorhabditis elegans
sp_plant:09LGM3	-	37.00	143.35	13.18	93	09LGM3 oryza sativa (rice). hypc
sp_rodent:09P9T9	-	36.00	132.24	22.66	225	09P9T9 mus musculus (mouse). 17
sp_human:096KZ8	+	36.00	132.17	22.66	227	096KZ8 homo sapiens (human). si
sp_invertebrate:09W1Z5	+	35.00	118.17	39.34	788	09W1Z5 drosophila melanogaster
sp_invertebrate:093636	+	35.00	111.62	40.22	1785	093636 caenorhabditis elegans
sp_invertebrate:093637	+	35.00	111.00	40.30	1929	093637 caenorhabditis elegans
sp_bacteriap:086710	+	34.00	125.86	63.47	182	086710 aquifex aeolicus. hypoth
sp_bacteriap:09A9E6	+	34.00	123.07	64.07	258	09A9E6 staphylococcus aureus (s
sp_bacteriap:09A9Z8	+	34.00	117.16	65.36	540	09A9Z8 caulobacter crescentus.
sp_invertebrate:09U0J6	+	34.00	104.67	68.17	2567	09U0J6 plasmodium falciparum
sp_human:09HAC7	-	33.00	114.87	109.05	434	09HAC7 homo sapiens (human). hy
sp_plant:09XJ26	-	33.00	114.11	109.33	477	09XJ26 nicotiana tabacum (commo
sp_bacteriap:087888	+	33.00	113.00	109.74	548	087888 lawsonia intracellularis
sp_bacteriap:092PG2	+	33.00	112.96	109.76	551	092PG2 rhizobium meliloti (sinc
sp_plant:0941B6	-	33.00	112.84	109.80	559	0941B6 nicotiana tabacum (commo
sp_bacteriap:098DS0	-	33.00	112.44	109.95	588	098DS0 rhizobium loti (mesorhiz
sp_invertebrate:09W0S9	+	33.00	105.80	112.44	1346	09W0S9 drosophila melanogaster
sp_invertebrate:09NGP2	+	33.00	104.22	113.04	1640	09NGP2 drosophila melanogaster
sp_bacteriap:09ETD0	+	32.00	120.06	177.43	137	09ETD0 corynebacterium equi (i
sp_virus:087706	-	32.00	116.95	179.30	202	087706 fowlpox virus (fpv). orf
sp_invertebrate:09GR98	+	32.00	116.20	179.76	222	09GR98 aphid gossypii (cotton
sp_human:09BVC5	-	32.00	115.84	179.97	232	09BVC5 homo sapiens (human). hy
sp_virus:090862	-	32.00	113.79	181.22	300	090862 yaba monkey tumor virus.
sp_bacteriap:098TY6	+	32.00	113.45	181.43	333	098TY6 rhizobium loti (mesorhiz
sp_vertebrate:09RH55	+	32.00	113.29	181.53	319	09RH55 gallus gallus (chicken).
sp_plant:09SRH3	+	32.00	113.20	181.59	323	09SRH3 arabidopsis thaliana (mc
sp_bacteriap:09AEU1	+	32.00	112.93	181.75	334	09AEU1 streptococcus gordonii.
sp_bacteriap:09PBK3	+	32.00	112.24	182.17	364	09PBK3 xylella fastidiosa. abc
sp_bacteriap:09CBM5	+	32.00	112.02	182.31	374	09CBM5 mycobacterium leprae. hy
sp_plant:09MAD5	+	32.00	111.94	182.36	378	09MAD5 arabidopsis thaliana (mc
sp_invertebrate:09GZF0	-	32.00	111.73	182.49	388	09GZF0 caenorhabditis elegans
sp_invertebrate:09XTH5	+	32.00	110.43	183.28	456	09XTH5 caenorhabditis elegans
sp_invertebrate:09TVL3	+	32.00	109.87	183.63	489	09TVL3 caenorhabditis elegans
sp_bacteriap:09CNX8	-	32.00	109.34	183.96	523	09CNX8 pasteurella multocida. g
sp_bacteriap:09TPS2	-	32.00	109.24	184.02	529	09TPS2 streptococcus pneumoniae
sp_plant:09LRF8	+	32.00	108.25	184.64	599	09LRF8 arabidopsis thaliana (mc
sp_plant:023223	+	32.00	107.97	184.81	620	023223 arabidopsis thaliana (mc
sp_plant:093Y36	+	32.00	107.77	184.94	636	093Y36 arabidopsis thaliana (mc

```
sp_fungi:09UTE5 + 32.00 107.73 184.96 639 09ute5 schizosaccharomyces p
sp_plant:09M9M5 + 32.00 107.73 184.96 639 09m9m5 arabidopsis thaliana
sp_plant:09ZPW5 + 32.00 106.97 185.44 703 09zpw5 arabidopsis thaliana
sp_invertebrate:093382 - 32.00 105.82 186.16 811 093382 caenorhabditis eleg
sp_bacteriap:09JTB6 + 32.00 104.56 186.95 950 09jtb6 neisseria meningitidi
```

seq\_name: sp\_invertebrate:018242

seq\_documentation\_block:

```
ID 018242 PRELIMINARY; PRT; 343 AA.
AC 018242;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Y57G11C.18 PROTEIN.
GN Y57G11C.18.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z99281; CAB16519.1; -.
SQ SEQUENCE 343 AA; 40942 MW; 617F6E8291BF77F8 CRC64;
```

alignment\_scores:

```
Quality: 38.00 Length: 8
Ratio: 4.750 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 75.000
```

alignment\_block:

```
US-09-730-559B-45/rev x 018242 ..
Align seg 1/1 to: 018242 from: 1 to: 343
```

```
24 CGATCTATCAATAATATTAAGCCC 1
:::|||||:::|||||
222 LysSerMetAsnVallysPro 229
```

seq\_name: sp\_invertebrate:016733

seq\_documentation\_block:

```
ID 016733 PRELIMINARY; PRT; 351 AA.
AC 016733;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE K09F6.1 PROTEIN.
GN K09F6.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
```

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,  
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RC STRAIN-BRISTOL N2;  
RP Scheet P.;  
RT "The sequence of C. elegans cosmid K09P6.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RT Waterston R.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF016683; AAB66192.1; -;  
SQ SEQUENCE 351 AA; 41772 MW; 8C0BBE35CAA25B15 CRC64;

alignment\_scores:  
Quality: 38.00 Length: 8  
Ratio: 4.750 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45/rev x 016733 ..

Align seg 1/1 to: 016733 from: 1 to: 351

24 CGATCTATGAATATATTAGCCC 1

263 LysSerMetAsnVallysPro 270

seq\_name: sp\_plant:Q9LGM3

seq\_documentation\_block:

ID Q9LGM3 PRELIMINARY; PRT; 93 AA.

AC Q9LGM3;

DT 01-OCT-2000 (TREMREL. 15, Created)

DT 01-OCT-2000 (TREMREL. 15, Last sequence update)

DT 01-MAR-2001 (TREMREL. 16, Last annotation update)

DE HYPOTHETICAL PROTEIN.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC

clone:P0041e1.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC

clone:P0433f09.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF002521; BAA96761.1; -;

DR EMBL; AP002539; BAB08200.1; -;

SQ SEQUENCE 93 AA; 10539 MW; CDBA1B7C77BB9D25 CRC64;

alignment\_scores:

Quality: 37.00 Length: 8

Ratio: 4.625 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 75.000

alignment\_block:

US-09-730-559B-45/rev x Q9LGM3 ..

Align seg 1/1 to: Q9LGM3 from: 1 to: 93

24 CGATCTATGAATATATTAGCCC 1

5 ArgSerIleAsnValIysPro 12

seq\_name: sp\_rodent:Q9D9T9

seq\_documentation\_block:

ID Q9D9T9 PRELIMINARY; PRT; 225 AA.

AC Q9D9T9;

DT 01-JUN-2001 (TREMREL. 17, Created)

DT 01-JUN-2001 (TREMREL. 17, Last sequence update)

DT 01-JUN-2001 (TREMREL. 17, Last annotation update)

DE 1700029F09RIK PROTEIN.

GN 1700029F09RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=TESTIS;

RX MEDLINE=21085660; PubMed=1117851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL; AK006487; BAB24612.1; -;

DR MGD; MGI:1922873; 1700029F09RIK.

DR InterPro; IPR000379; Est\_lip\_thioest\_actsite.

SQ SEQUENCE 225 AA; 25240 MW; F713B5D975EF3249 CRC64;

alignment\_scores:

Quality: 36.00 Length: 7

Ratio: 5.143 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:

US-09-730-559B-45 x Q9D9T9 ..

Align seg 1/1 to: Q9D9T9 from: 1 to: 225

1 GGCCTAATATTATTCATAGA 21

|||||||||||||:|||||

68 GlyLeuAsnIleValHisarg 74

seq\_name: sp\_human:Q96KZ8

seq\_documentation\_block:

ID Q96KZ8 PRELIMINARY; PRT; 227 AA.

AC Q96KZ8;



DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE SIMILAR TO RIKEN CDNA 1700029F09 GENE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Strausberg R.;  
RL Submitted (RC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC015148; AAI15148.1; -  
SQ SEQUENCE 227 AA; 25585 MW; B37C06A38A91E774 CRC64;

alignment\_scores:  
Quality: 36.00 Length: 7  
Ratio: 5.143 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

## alignment\_block:

US-09-730-559B-45 x Q96KZ8 ..  
Align seg 1/1 to: Q96KZ8 from: 1 to: 227

1 GGGCTTAATATTATTCATAGA 21  
|||||  
68 GlyLeuAsnIleValHisArg 74

seq\_name: sp\_invertebrate:Q9W1Z5

## seq\_documentation\_block:

ID Q9W1Z5 PRELIMINARY; PRT; 788 AA.  
AC Q9W1Z5;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CG13528 PROTEIN.  
GN CG13528.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G.G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agyaniani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., P.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siding-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003458; AAF46906.1; -.  
DR FlyBase; FBgn0034779; CG13528.  
DR InterPro; IPR002110; ANK.  
DR Pfam; PF00023; ank; 4.  
DR SMART; SM00248; ANK; 3.  
DR PROSITE; PS50088; ANK\_REPEAT; 3.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 2.  
KW ANK repeat; Repeat.  
SQ SEQUENCE 788 AA; 85586 MW; 5E03F8124B07474C CRC64;

alignment\_scores:  
Quality: 35.00 Length: 8  
Ratio: 4.375 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.000

## alignment\_block:

US-09-730-559B-45 x Q9W1Z5 ..  
Align seg 1/1 to: Q9W1Z5 from: 1 to: 788

1 GGGCTTAATATTATTCATAGTCG 24  
|||||  
118 GlyLeuAsnIleHisLysAla 125

seq\_name: sp\_invertebrate:Q93636

## seq\_documentation\_block:

ID Q93636 PRELIMINARY; PRT; 1785 AA.  
AC Q93636;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE F29G6.3B PROTEIN.  
GN F29G6.3B.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Harris B.R.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; Z78543; CAB01754.1; -.  
DR InterPro; IPE004019; YLP.  
DR Pfam; PF02757; YLP; 8.  
SQ SEQUENCE 1785 AA; 205049 MW; F92FFC5ABEC286CD CRC64;

alignment\_scores:  
Quality: 35.00 Length: 8

```
Ratio: 4.375          Gaps: 0
Percent Similarity: 100.000    Percent Identity: 62.500
alignment_block:
US-09-730-559B-45 x Q93636    ..
Align seg 1/1 to: Q93636 from: 1 to: 1785

1 GGCCTTAATATTATTCATAGATCG 24
||||:||||:||||:||||:||||:
261 GlyValAsnValLeuHisArgSer 268

seq_name: sp_invertebrate:Q93637

seq_documentation_block:
ID Q93637 PRELIMINARY; PRT; 1929 AA.
AC Q93637;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F29G6.3A PROTEIN.
GN F29G6.3A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; 278543; CAB01755.1; -.
SQ SEQUENCE 1929 AA; 220417 MW; C45166E4AAACE23B9 CRC64;

alignment_scores:
Quality: 35.00          Length: 8
Ratio: 4.375           Gaps: 0
Percent Similarity: 100.000    Percent Identity: 62.500
alignment_block:
US-09-730-559B-45 x Q93637    ..
Align seg 1/1 to: Q93637 from: 1 to: 1929

1 GGCCTTAATATTATTCATAGATCG 24
||||:||||:||||:||||:||||:
261 GlyValAsnValLeuHisArgSer 268

seq_name: sp_bacteriap:O66710

seq_documentation_block:
ID O66710 PRELIMINARY; PRT; 182 AA.
AC O66710;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 19.9 KDA PROTEIN.
GN AQ_389.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RL MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
```

```
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
DR EMBL; AE000888; AAC06687.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 182 AA; 19894 MW; DEF9E5FD6594D074 CRC64;

alignment_scores:
Quality: 34.00          Length: 7
Ratio: 4.857           Gaps: 0
Percent Similarity: 100.000    Percent Identity: 85.714
alignment_block:
US-09-730-559B-45 x O66710    ..
Align seg 1/1 to: O66710 from: 1 to: 182

1 GGCCTTAATATTATTCATAGA 21
||||:||||:||||:||||:
43 GlyLeuAsnIleHisLys 49

seq_name: sp_bacteriap:Q99RE6

seq_documentation_block:
ID Q99RE6 PRELIMINARY; PRT; 258 AA.
AC Q99RE6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SA2275 (HYPOTHETICAL PROTEIN SAV2487).
GN SA2275 OR SAV2487.
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879; 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43578.1; -.
DR EMBL; AP003365; BAB58649.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 258 AA; 30455 MW; CCF5B0684B2D33BD CRC64;

alignment_scores:
Quality: 34.00          Length: 8
Ratio: 4.250           Gaps: 0
Percent Similarity: 100.000    Percent Identity: 62.500
alignment_block:
US-09-730-559B-45 x Q99RE6    ..
Align seg 1/1 to: Q99RE6 from: 1 to: 258

2 GGCCTTAATATTATTCATAGATCGA 25
||||:||||:||||:||||:
94 GlyLeuValLeuTyrValAspArg 101
```

seq\_name: sp\_bacteriap:Q9A928

seq\_documentation\_block:

ID Q9A928 PRELIMINARY; PRT; 540 AA.  
 AC Q9A928;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN CC0813.  
 GN CC0813.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.B., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AF005758; AAK22798.1; -.  
 DR TIGR; CC0813; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 540 AA; 59648 MW; 72BC4542BEF99FD CRC64;

alignment\_scores:  
 Quality: 34.00 Length: 8  
 Ratio: 4.857 Gaps: 0  
 Percent Similarity: 87.500 Percent Identity: 87.500

alignment\_block:

US-09-730-559B-45 x Q9A928 ..

Align seg 1/1 to: Q9A928 from: 1 to: 540

2 GGCTTAATATATTCATAGATCGA 25  
 |||||  
 431 GlyLeulleuPheTyAspArg 438

seq\_name: sp\_invertebrate:Q9U0J6

seq\_documentation\_block:

ID Q9U0J6 PRELIMINARY; PRT; 2567 AA.  
 AC Q9U0J6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL 312.5 KDA PROTEIN.  
 GN MAL4P2.26.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,  
 RA Quail M., Barrell B.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035475; CAB62867.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2567 AA; 312499 MW; F58BA2994C211F47 CRC64;

alignment\_scores:  
 Quality: 34.00 Length: 8

Ratio: 4.250 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-09-730-559B-45 x Q9U0J6 ..

Align seg 1/1 to: Q9U0J6 from: 1 to: 2567

1 GGGCTTAATATATTCATAGATCG 24  
 |||||  
 932 GlyLeuAsnValIleHisLysAsn 939

seq\_name: sp\_human:Q9HAC7

seq\_documentation\_block:

ID Q9HAC7 PRELIMINARY; PRT; 434 AA.  
 AC Q9HAC7;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE HYPOTHETICAL 47.2 KDA PROTEIN.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masubo Y., Oshima A.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK021870; BAB13922.1; -.  
 DR InterPro; IPR003673; CAIB-BAIF.  
 DR Pfam; PF02515; CAIB-BAIF; 1.  
 DR PF02515; CAIB-BAIF; 1.  
 SQ SEQUENCE 434 AA; 47186 MW; 7154D3E23618A883 CRC64;

alignment\_scores:

Quality: 33.00 Length: 6  
 Ratio: 5.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-730-559B-45/rev x Q9HAC7 ..

Align seg 1/1 to: Q9HAC7 from: 1 to: 434

18 ATCAATAATATTAGCCC 1  
 |||||  
 42 MetAsnIleLysPro 47

seq\_name: sp\_plant:Q9XJ26

seq\_documentation\_block:

ID Q9XJ26 PRELIMINARY; PRT; 477 AA.  
 AC Q9XJ26;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NTWRKY1.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. XANTHI;  
 RA Shinshi H., Yamamoto S., Suzuki K.;

RT "Analyses of an elicitor-responsive element and transcription factors  
RT in cultured tobacco cells."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB022693; BAA82107.1; -.  
DR InterPro: IPR003657; WRKY.  
DR Pfam: PF03106; WRKY; 2  
SQ SEQUENCE 477 AA; 53053 MW; 24B8617D8C59B7A9 CRC64;

alignment\_scores:  
Quality: 33.00 Length: 6  
Ratio: 5.500 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-730-559B-45/rev x Q9XJ26 ..  
Align seg 1/1 to: Q9XJ26 from: 1 to: 477

18 ATGAAATATATTAGCCC 1  
|||||  
456 MetAsnAileIelysPro 461

seq\_name: sp\_bacteria:O87888

seq\_documentation\_block:

ID O87888 PRELIMINARY; PRT; 548 AA.  
AC O87888;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN).  
OS Lawsonia intracellularis.  
OG Plasmid pISI-2.  
OC Bacteria; Proteobacteria; delta subdivision; Lawsonia.  
OX NCBI\_TaxID=29546;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98386497; Pubmed=9720028;  
RA Dale C.J.H., Moses E.K., Ong C.C., Morrow C.J., Reed M.B., Hasse D.,  
RA Strugnelli R.A.;  
RT "Identification and sequencing of the groE operon and flanking genes  
of Lawsonia intracellularis: use in phylogeny."  
RL Microbiology 144:0-0(0).  
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND  
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS  
CC CONDITIONS (BY SIMILARITY).  
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF  
CC 7 SUBUNITS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.  
DR EMBL; U45241; AAC36500.1; -.  
DR HSP; P06139; IGRU.  
DR InterPro: IPR001844; Chaperonins\_cpn60.  
DR InterPro: IPR002423; TCPl\_cpn60.  
DR Pfam: PF00118; cpn60\_TCPl; 1.  
DR PRINTS; PR00298; CHAPERONIN60.  
DR PRINTS; PR00304; TCOMPLEXTCPl.  
DR PROSITE; PS00296; CHAPERONIN\_CPN60; 1.  
KW ATP-binding; Chaperone; Plasmid.  
SQ SEQUENCE 548 AA; 58605 MW; 6388C431E663E498 CRC64;

alignment\_scores:  
Quality: 33.00 Length: 8  
Ratio: 4.125 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:

US-09-730-559B-45 x O87888 ..  
Align seg 1/1 to: O87888 from: 1 to: 548

1 GGGCTTAATATTATCATGATCG 24

|||||  
439 GlyLeuAsnIleIleArgArgSer 446



